

Financial market volatility : statistical models and empirical analysis

Citation for published version (APA):

Mahieu, R. J. (1995). *Financial market volatility : statistical models and empirical analysis*. [Doctoral Thesis, Maastricht University]. Datawyse / Universitaire Pers Maastricht.
<https://doi.org/10.26481/dis.19951124rm>

Document status and date:

Published: 01/01/1995

DOI:

[10.26481/dis.19951124rm](https://doi.org/10.26481/dis.19951124rm)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Financial Market Volatility

Statistical Models and Empirical Analysis

Financial Market Volatility

Statistical Models and Empirical Analysis

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de
Rijksuniversiteit Limburg te Maastricht,
op gezag van de Rector Magnificus, Prof. Mr. M.J. Cohen,
volgens het besluit van het College van Dekanen,
in het openbaar te verdedigen op vrijdag 24 november 1995 om 14.00 uur

door

Ronaldus Johannes Mahieu

Promotores :

Prof. dr. P.C. Schotman

Prof. dr. C.C.P. Wolff

Beoordelingscommissie :

Prof. dr. C.G. Koedijk (voorzitter)

Prof. dr. Th.E. Nijman (Katholieke Universiteit Brabant)

Prof. dr. F.C. Palm

Aan mijn ouders

Preface

The charm of any research project undertaken is that it raises more questions than it can answer. This is also the experience that I had with the research of which this thesis is the result. I feel that the study of finance is still in its infancy and although much research has already been done, the number of unresolved issues remains large.

Research is an activity that should not be undertaken alone. During the work on this thesis and other scientific projects, I have had the support of many people. First of all, I would like to thank my thesis supervisors Peter Schotman and Christian Wolff. They provided me with invaluable comments and criticisms on many issues during my research. Also, I would like to thank the people that were involved, in one way or the other, in the writing of the papers on which a number of chapters of this thesis were based: Rob Bauer, Piet Eichholtz, Kees Koedijk, and Fred Nieuwland. Furthermore, I thank the members of the finance group at the University of Limburg for providing a creative and productive atmosphere.

During my stay in Maastricht, I have had the opportunity to visit many scientific conferences to present some of my work. The Netherlands Organization for the Advancement of Pure Research (NWO) is greatly acknowledged for giving me financial support. I also thank the Woodrow Wilson School of International and Public Affairs at Princeton University for hosting me as a visiting student for five months. The ABP is acknowledged for letting me get a glimpse of "real-life" finance. A word of special thanks goes to Stefano Cavaglia who introduced me to the London Parallel Applications Centre (LPAC), where many of the computations in this thesis were performed. I thank Peter Dzwig and Adriaan Joubert of LPAC for generously offering me office space and other facilities.

Without the moral support of a number of people this thesis could not have been completed. First of all, I thank my parents, brother and sister for giving me all the love and support that is within their capabilities. Friends time and time again made it clear to me that

research is not the only thing in life. And surely it is not.

Finally, I want to express my appreciation and gratitude to Marian, for putting up with my divided attention during this book's gestation period.

Ronald Mahieu,
August, 1995

Contents

| | |
|---|------------|
| Preface | vii |
| List of Tables | xii |
| List of Figures | xiv |
| 1 Introduction | 1 |
| 1.1 Outline | 5 |
| 2 Volatility Models | 9 |
| 2.1 Conditional Heteroskedasticity Models | 10 |
| 2.2 A Unified Framework of Volatility | 12 |
| 2.3 Stochastic Volatility Models | 13 |
| 2.4 Moment Comparison | 16 |
| 2.5 Conclusions | 18 |
| 3 The Estimation of Stochastic Volatility Models | 19 |
| 3.1 Estimation Techniques | 21 |
| 3.1.1 Non-Simulation Estimators | 22 |
| 3.1.2 Simulation Estimators | 27 |
| 3.2 Mixture Model | 30 |
| 3.3 Parameter Estimates | 37 |
| 3.4 Volatility Estimates | 47 |
| 3.5 Conclusions | 53 |
| Appendix 3.A: Bayesian Simulation Algorithm | 54 |
| Appendix 3.B: Moments | 56 |
| 4 Neglected Common Factors in Exchange Rate Volatility | 59 |
| 4.1 A Factor Structure of Exchange Rate News | 60 |
| 4.2 Multivariate Time Varying Volatility | 67 |
| 4.2.1 Specification Problems | 67 |

| | | |
|----------|--|------------|
| 4.2.2 | A Multivariate Stochastic Volatility Model for Exchange Rates | 68 |
| 4.3 | Stochastic Volatility Results | 70 |
| 4.4 | Implications for Risk Premia | 77 |
| 4.5 | Conclusions | 84 |
| 5 | Volatility and Volume: Balancing the Scales | 85 |
| 5.1 | The Mixture of Distributions Hypothesis and Stochastic Volatility Models . | 87 |
| 5.2 | Data Description and Leverage Effects | 89 |
| 5.3 | Estimation Results | 97 |
| 5.4 | Conclusions | 106 |
| | Appendix 5: VAR Analysis | 107 |
| 6 | Hedging Foreign Currency Exposure: Do Fat Tails Matter? | 119 |
| 6.1 | Currency Hedging Revisited | 121 |
| 6.2 | Optimal Hedge Ratios | 128 |
| 6.3 | Conclusions | 138 |
| 7 | A Factor Approximation Model for the International Arbitrage Pricing Theory | 139 |
| 7.1 | International Arbitrage Pricing Theory | 140 |
| 7.2 | Factor Approximation | 144 |
| 7.3 | Conclusions | 147 |
| 8 | Conclusions | 149 |
| | Bibliography | 155 |
| | Nederlandse samenvatting | 163 |
| | Index | 168 |
| | Curriculum Vitae | 171 |

List of Tables

| | | |
|------|---|-----|
| 3.1 | Mixture Parameters | 35 |
| 3.2 | Summary statistics of $y_t = \ln[(\Delta \ln S_t - \bar{s})^2]$ | 38 |
| 3.3 | Parameter Estimates of Stochastic Volatility Model | 41 |
| 3.4 | Bayesian Posterior Moments | 43 |
| 3.5 | Likelihood Ratio Tests | 43 |
| 3.6 | Moments of Measurement Error Distribution | 44 |
| 3.7 | Data and Implied Model Moments | 45 |
| 3.8 | Diagnostic LM tests | 47 |
| 3.9 | Summary Statistics of Volatility Estimates | 52 |
| 4.1 | Variance Decomposition of Exchange Rate News | 64 |
| 4.2 | Sample and Fitted Moments of s_{ij} | 65 |
| 4.3 | Diagnostics of News Components | 66 |
| 4.4 | Summary Statistics of Transformed Exchange Rate Changes $w_{ij}(t)$ | 71 |
| 4.5 | Parameter Estimates of Stochastic Volatility Model | 73 |
| 4.6 | Diagnostics of Stochastic Volatility Model | 74 |
| 4.7 | Stochastic Volatility Tests | 77 |
| 4.8 | Summary Statistics of Risk Premia | 83 |
| 5.1 | Summary Statistics Adjusted Stock Returns | 91 |
| 5.2 | Summary Statistics Adjusted Log Volume | 92 |
| 5.3 | Sign Bias Tests | 94 |
| 5.4 | Leverage Regressions | 96 |
| 5.5 | Stochastic Volatility Estimation Results I: | 98 |
| 5.6 | Stochastic Volatility Estimation Results II: | 99 |
| 5.7 | Stochastic Volatility Estimation Results III: | 100 |
| 5.8 | Stochastic Volatility Estimation Results IV: | 101 |
| 5.9 | Optimal VAR Orders | 109 |
| 5.10 | Description Impulse Response Functions Trivariate VAR | 110 |
| 5.11 | Granger-Causality Tests on Trivariate VAR | 117 |

6.1 Descriptive Statistics Stock Index Returns and Exchange Rate Returns . . 124

6.2 Correlation Matrices 125

6.3 Portfolio Decompositions for 3 Bilateral Portfolios 126

6.4 Portfolio Characteristics of Hedged Portfolios r_{t+1}^h 133

6.5 Hedged Portfolios: Student- t 137

List of Figures

| | | |
|------|---|-----|
| 3.1 | Probability Density Functions of ϵ_t (A) and $\xi_t = \ln \epsilon_t^2$ (B). | 24 |
| 3.2 | Student- t Densities of ϵ_t (A) and $\xi_t = \ln \epsilon_t^2$ (B). | 25 |
| 3.3 | Densities of Measurement Error | 35 |
| 3.4 | Posterior Densities of ϕ | 40 |
| 3.5 | Implied Measurement Error Densities | 42 |
| 3.6 | Estimated Volatilities: yen/dollar | 48 |
| 3.7 | Estimated Volatilities: mark/dollar | 48 |
| 3.8 | Estimated Volatilities: pound/dollar | 49 |
| 3.9 | Estimated Volatilities: mark/yen | 49 |
| 3.10 | Estimated Volatilities: pound/yen | 50 |
| 3.11 | Estimated Volatilities: mark/pound | 50 |
| 4.1 | Time Series of Estimated Conditional Log-Variances | 75 |
| 4.2 | Time Series of "Smoothed" Estimates of the Log-Volatilities | 76 |
| 4.3 | Conditional Variance of yen/dollar Exchange Rate | 79 |
| 4.4 | Conditional Variance of mark/dollar Exchange Rate | 79 |
| 4.5 | Conditional Variance of pound/dollar Exchange Rate | 80 |
| 4.6 | Conditional Variance of mark/yen Exchange Rate | 80 |
| 4.7 | Conditional Variance of pound/yen Exchange Rate | 81 |
| 4.8 | Conditional Variance of pound/mark Exchange Rate | 81 |
| 5.1 | Hewlett-Packard: Returns | 90 |
| 5.2 | Hewlett-Packard: Trading Volumes (logarithms) | 90 |
| 5.3 | Volatilities Alcoa | 103 |
| 5.4 | Volatilities Amoco | 103 |
| 5.5 | Volatilities Bristol-Myers | 104 |
| 5.6 | Volatilities Champion | 104 |
| 5.7 | Volatilities Hewlett-Packard | 105 |
| 5.8 | Volatilities S&P 500 | 105 |
| 5.9 | Impulse Responses Trivariate VAR: Alcoa | 111 |

| | | |
|------|--|-----|
| 5.10 | Impulse Responses Trivariate VAR: Amoco | 112 |
| 5.11 | Impulse Responses Trivariate VAR: Bristol-Myers | 113 |
| 5.12 | Impulse Responses Trivariate VAR: Champion | 114 |
| 5.13 | Impulse Responses Trivariate VAR: Hewlett-Packard | 115 |
| 5.14 | Impulse Responses Trivariate VAR: S&P 500 | 116 |
| 6.1 | Means and Standard Deviations of r_{t+1}^h | 131 |
| 6.2 | Standard Deviations and Kurtosis of r_{t+1}^h | 132 |
| 6.3 | Expected and Sampled Frequencies of Fully Hedged Returns | 136 |

Chapter 1

Introduction

In the last few decades the study of financial economics has expanded enormously, both from a theoretical and from an empirical viewpoint. The interest in this field of research has not remained within the walls of academia though. Practitioners have taken up the academic theories and have applied them successfully on the trading floors all over the world. A productive cross-fertilization has emerged in which academic researchers and practitioners have pushed up the pace of developing new theories and refining old ones. The ongoing process of financial innovation that started in the 1970s, is tightly connected to developments in financial research. Empirical phenomena from financial markets are conscientiously analyzed by theorists. New techniques, often based on advanced mathematical concepts, find their way into trading practice at an ever increasing rate. Likewise, the surge in computer power has given applied financial research an enormous injection. Valuation of complicated financial assets can be performed in the blink of an eye. Also, electronic trading systems have facilitated the expansion of trading on financial markets even more.

The hectic scene sketched above contrasts sharply with the era of the founders of investment theory. When Markowitz and Roy independently developed the theory of portfolio selection in the 1950s financial markets were in a quiet era and quantitative techniques were not used at all. For a long time practitioners were reluctant to apply the techniques of portfolio selection (Bernstein, 1992). The ideas of Markowitz and Roy are based on a very general principle in economics: to obtain something we have to abandon something else. In the financial context a trade-off exists between expected return and risk. If an investor wants a higher expected return, he has to accept a higher risk as well. The theories of Markowitz and Roy – and all the other financial theories that have been developed since – deal with the quantification of the risk-return trade-off.

The notion of risk is very important in economics. In common parlance risk is often put on one level with uncertainty. From a methodological viewpoint, however, this is not correct. The treatise of Knight (1921) gives some insights in the matter of risk and its relation to uncertainty. Basically, risk is measurable with respect to some benchmark while uncertainty is not. That is, probabilities can be assigned to a number of possible outcomes in the case of risk. Originally set up as an analysis of entrepreneurship Knight's exposition tries to give a rationale for the emergence of profit in a market with approximate perfect competition. He argues that uncertainty, and not risk, is responsible for the existence of entrepreneurial profits. People can insure themselves against risk, because it is perfectly measurable. Knight illuminates this with an example of a bottling plant:

..., the bursting of bottles does not introduce an uncertainty or hazard into the business of producing champagne; since in the operations of any producer a practically constant and known proportion of the bottles will burst, it does not especially matter even whether the proportion is large or small. The loss becomes a fixed cost in the industry and is passed on to the consumer, like the outlays for labor or materials or any other. And even if a single producer does not deal with a sufficiently large number of cases of the contingency in question (in a sufficiently short period of time) to secure constancy in its effects, the same result may easily be realized, through an organization taking in a large number of producers. This, of course, is the principle of insurance, ... (Knight, 1921, p. 213).

In order for insurance to work, we need some kind of method to infer the probabilities of the contingencies. Knight discusses two ways. In the first place he introduces *a priori* or mathematical probabilities, which are obtained deductively. An example is the probability distribution of throwing a perfect dice. We do not have to perform an experiment in order to find out the actual probabilities. Even if we did, the outcomes will be looked upon with suspicion from a mathematical side. The second kind of probability is more relevant for insurance. It relies exclusively on past experience, and is consequently called statistical probability (Feller, 1968). The statistical attitude toward probability has been developed mainly by Fisher and Von Mises. As opposed to mathematical probability the statistical probability of a certain event occurring cannot be given accurately. For example, it is impossible to give the probability that a particular building will burn down. In fact, this probability itself is a random variable.

Uncertainty can be put in a probability context as well. However, it differs from the mathematical and statistical types of probabilities in one important aspect. The proba-

bility statements made under uncertainty have no valid basis of any kind for classifying instances. In other words, probability statements are purely subjective. For example, the probability of a meteorite impact in the Netherlands cannot be given by mathematical or statistical probabilities. The notion of classification is of major importance here. If we could make a classification of possible outcomes the probability statements we make become less subjective and, in a way, become more certain. In many real-life situations classifications cannot be made and we are dealing with true uncertainty. Knight notes the following with regard to making business decisions under uncertainty:

Yet it is true, and the fact can hardly be over-emphasized, that a judgement of probability is actually made in such cases. The business man himself not merely forms the best estimate he can of the outcome of his actions, but he is likely also to estimate the probability that his estimate is correct. The "degree" of certainty or of confidence felt in the conclusion after it is reached cannot be ignored, for it is of the greatest practical significance. The action which follows upon an opinion depends as much upon the amount of confidence in that opinion as it does upon the favorableness of the opinion itself. The ultimate logic, or psychology, of these deliberations is obscure, a part of the scientifically unfathomable mystery of life and mind. (Knight, 1921, pp. 226–227).

In financial economics risk is usually defined to be a property of a set of random outcomes. Typically, these outcomes are determined by past experience and consequently can be regarded as statistical probabilities. No explicit account is taken for uncertainty, *i.e.* the information on which an investor bases his decisions is determined by a – possibly infinite – set of random variables only (Ingersoll, 1987). The random variables take on different values for each state of the world that the investor perceives to be attainable. Of course, this does not mean that uncertainty is not present anymore, only that it does not enter the model. Modelling uncertainty is a sheer impossibility, because no classification of states of the world can be made. Knight's theories predict that in a frictionless world no profits can arise in a world that is risky, but not uncertain. It should be remembered that the definition of profit is tightly connected to the amount of uncertainty. We can transfer this notion directly to a financial framework. In a perfect world in which markets are efficient, investors receive returns in accordance with the amount of risk they have taken upon them and no "entrepreneurial" profits will be made. In an imperfect world uncertainties will be present and profit opportunities arise. For instance, markets may be inefficient and investors do not have enough information to exclude uncertainty from their decisions. In such an environment profits can be made, *e.g.* as a result from arbitrage. In that case the

"entrepreneurship" of the arbitrageur is rewarded by an arbitrage profit.

Usually, the investor summarizes his perceptions on the relevant random variables in an objective function. By giving appropriate weights to the possible outcomes of every random variable the investor is able to adjust his perceptions towards risk. In financial theory the device for doing this is a utility function. An investor that dislikes risk will have a utility function that downplays the probability outcomes that are perceived to be risky. In other words, the investor shapes his utility function according to his attitude towards risk. A measure of risk that springs to mind immediately is the dispersion of possible outcomes. If the outcomes of a particular random variable (*e.g.* a return on a single asset or a portfolio of assets) can take extreme values with high probability, both positive and negative, relative to another random variable, the former is disliked by the investor. However, the choice whether or not the first random variable is preferred to the second depends not only on risk. The expected value (return) of the variables and possibly higher moments (skewness, kurtosis) can be important as well. The utility function of the investor takes all these probability characteristics into account and allows the investor to make the "right" decisions.

In financial applications risk is often associated with variance or standard deviation. Using a theorem of Rothschild and Stiglitz (1970) we can show that for concave utility functions and general distribution functions the variance of a more risky variable should be bigger. The theorem is

Theorem: Outcome \tilde{X} is less risky than outcome \tilde{Y} if and only if \tilde{Y} is distributed like $\tilde{X} + \tilde{\epsilon}$ and $\tilde{\epsilon}$ is a fair game with respect to \tilde{X} . In other words

$$E[\tilde{\epsilon}|X] = 0, \quad \text{for all } X$$

A proof of this theorem can be found in Ingersoll (1987). An important remark that can be made as a result of this theorem is that if \tilde{Y} is riskier than \tilde{X} , then $\text{Var}(\tilde{Y}) \geq \text{Var}(\tilde{X})$. This can be seen as follows. The condition $E[\tilde{\epsilon}|X] = 0$ implies that $\text{Cov}(\tilde{X}, \tilde{\epsilon}) = 0$. Then

$$\begin{aligned} \text{Var}(\tilde{Y}) &= \text{Var}(\tilde{X} + \tilde{\epsilon}) \\ &= \text{Var}(\tilde{X}) + \text{Var}(\tilde{\epsilon}) + 2\text{Cov}(\tilde{X}, \tilde{\epsilon}) \\ &= \text{Var}(\tilde{X}) + \text{Var}(\tilde{\epsilon}) \\ &\geq \text{Var}(\tilde{X}) \end{aligned}$$

In case of normally distributed random variables variance and risk are equivalent. Variance has played an important role in financial decision making. Variances (and means) of returns

are crucial in the derivation of the Capital Asset Pricing Model. A large body of literature in financial economics has investigated the properties of financial time series. One of the main conclusions that arises from these studies is that financial time series show a considerable amount of leptokurtosis, *i.e.* the distributions are characterized by more mass in the tails and around zero than is the case with normal distributions. This implies that risk measures cannot be based on the variance solely. However, Levy and Markowitz (1979) have shown that even under these circumstances expected utility maximization is predominantly characterized by mean and variance of the distribution.

In the discussion up to now, we have concentrated on risk and its relation to variance. Variance is a statistic that can be determined from a past history of prices. In most financial models variance is used as a proxy for the present *variability* of returns. Variability is the quantity that we are really interested in. The question that we might ask at this moment relates to the validity of using (sample) variance to say something about today's variability. Intuitively, we want a measure for variability that conforms as closely to the current variability as possible. For this we need a broader concept. The underlying process of variability is called *volatility*. Andersen (1992) gives a classification of different volatility models on the basis of measurability concepts. That is, the type of variability that each volatility model describes is determined by the maximum amount of information that is allowed under a specific probability measure.

In this thesis we will present and discuss several volatility models. The models will be confronted with data series in order to check their abilities in describing volatility. Volatility models abound in the financial literature. The advent of financial futures and other derivative securities on financial markets (Miller, 1991) has given an enormous boost to research on volatility. A very important role in the valuation procedures of these securities is awarded to volatility. The enormous proliferation in financial trading on the world's financial market places that we have witnessed in recent years can largely be attributed to the development and use of derivative securities. The pace at which all these developments takes place is still growing at an exponential rate. A better understanding of volatility processes might preclude the recent disasters with derivative securities in financial practice.

1.1 Outline

This thesis consists of the following chapters. In Chapter 2 we will introduce the concept of volatility in a more rigorous way. Two important types of volatility models will be

discussed and compared. The first is the AutoRegressive Conditional Heteroskedasticity (ARCH) class of models, that has been developed by Engle (1982). This article has led to a plethora of other volatility models of which Bollerslev's (1986) Generalized ARCH (GARCH) and Nelson's (1991) Exponential GARCH (EGARCH) are the most notable. Spurred by the empirical success of ARCH models several researchers have begun to develop other volatility models. A special type of models consists of the so-called Stochastic Volatility (SV) models, which, in some sense, could be regarded as extensions to ARCH. We will discuss the theoretical differences between SV and ARCH models within a general framework of volatility. SV models are natural empirical equivalents to the continuous-time models that are widely used in theoretical finance. The application of SV models to problems in financial economics is one of the main themes in this thesis.

The estimation of SV models is not a trivial task. Maximum Likelihood estimation cannot be performed in a straightforward matter as is the case for models from the ARCH class. In Chapter 3 we will discuss the problems of estimating univariate SV models. Quasi Maximum Likelihood methods and the (Generalized) Method of Moments have been used extensively in empirical analyses, but compared to other estimation procedures these methods are inefficient. A large part of this chapter will be devoted to simulation-type estimators of SV models. These estimators, which become more and more popular in econometrics, give more efficient estimates. Application of complicated models is facilitated by the ongoing surge in computer power. We will estimate the SV model by simulation techniques that are based on the Gibbs sampler. We use both a Bayesian as well as a likelihood approach to estimate the parameters of the model. We analyze the models by comparing parameter and volatility estimates using a sample of weekly exchange rate data for four major currencies over the floating-rate period.

The specification of multivariate volatility models has always been a difficult task in finance. Straightforward extensions of univariate ARCH models to the multivariate case are not feasible for even a limited number of time series. This is mainly due to the proliferation of parameters that need to be estimated. Several attempts have been made to keep the number of parameters limited within multivariate ARCH models. The most successful models that have been presented in the literature exploit some kind of factor structure on a set of time series. Given the success of Ross's (1976) Arbitrage Pricing Theory (APT), this seems to be a sensible way of modelling returns. In Chapter 4 we will present a factor model for exchange rate changes with multivariate stochastic volatility. The model is independent of the numeraire currency of the system of exchange rates. That is, the results remain unchanged if we switch the currency of denomination from, for example, the U.S. dollar to the German mark.

In Chapter 5 we will analyze the relationship that exists between stock return volatility and its corresponding trading volume. Several theories predict that volumes should have a significant information content with respect to volatility. This line of research has tried to give statistical models for volatility some theoretical background. In financial markets volatility and volume are determined simultaneously. This leads us to consider trading volume as a scale variable for volatility in Chapter 5. We perform our analyses on a long series of daily observations on five individual stocks that are traded on the New York Stock Exchange as well as on the Standard & Poors 500 index. We also performed some analyses on the existence of the so-called leverage effect. The leverage effect predicts that shocks in volatility have asymmetric effects on the returns of the stock. The empirical analyses in this chapter rely heavily on the (univariate) SV model again. In the appendix of this chapter we will explore the dynamic relations between volume, volatility and returns by estimating some Vector Autoregressive (VAR) models.

Volatility in exchange rates plays a major role in international portfolio diversification. Returns on foreign investments can be adversely affected by exchange rate movements. Therefore, investors are interested in ways to diminish exchange rate risk on their international portfolios. In Chapter 6 the topic of hedging exchange rate exposure is taken up. We develop a framework in which currency hedging depends on the of risk aversion of the investor as well as on the leptokurtosis of a portfolio consisting of domestic and foreign investments, exchange rates and forward contracts. We assume a fat-tailed distribution on the total portfolio return and compute optimal hedge ratios by optimizing an expected utility function.

In Chapter 7 we present a theoretical extension to arbitrage pricing models in an international setting. In finance arbitrage pricing models try to capture the co-movements in a large set of returns by a small amount of factors. The existing models on international portfolio diversification are not very flexible with respect to the desired numeraire independence of an international asset pricing model. We need the model to be invariant with respect to the numeraire currency, because otherwise investors from different countries could profit from arbitrage. We present a model that preserves the numeraire independence and at the same time allows a flexible factor structure for asset and exchange rate returns. Our exposition hinges on the result that under specific circumstances factors in a factor model can be approximated by principal components. Since, by construction, a principal components analysis preserves the variance-covariance structure of the return series, it is possible to obtain a rotation on the components that makes the approximate factors numeraire-invariant. Consequently, some restrictions on international APT models, that were needed in earlier studies, can be relaxed.

In Chapter 8 we will end this thesis with a summary and some concluding remarks.

Chapter 2

Volatility Models

From the previous chapter we have learned that volatility should be an important ingredient in various kinds of financial decision making processes. More specifically, in topics like portfolio allocation, contingent claims pricing and hedging, volatility plays a crucial role. In this chapter we will introduce two types of volatility models. First, we will discuss models of the AutoRegressive Conditional Heteroskedasticity (ARCH) class. The research of these models has been initiated by Engle (1982). Basically, ARCH models describe how conditional volatility changes over time as a function of information from the past. An alternative approach, which has become more popular recently, is the Stochastic Volatility (SV) model, where the volatility is modelled as an unobserved component that follows some stochastic process. SV models have been proposed because they relate directly to the type of diffusion processes used in asset pricing theory in theoretical finance. Initial research in these directions was performed by Clark (1973). Although the theoretical merits of SV models were well understood, empirical analysis started only a few years ago.

The two types of volatility models described above are by no means the only volatility models that have been used in the literature. Bollerslev, Engle and Nelson (1993) give a short overview of other approaches to describe time-varying second moments. These include the estimation of instantaneous (implied) volatility through inversion of the well-known Black-Scholes formula for valuing options, which gives market-based estimates of volatility.

Both ARCH and SV models are statistical models. No satisfactory structural dynamic economic theory exists for the explanation of time variation in higher moments. Attempts have been made from the literature on the mixture of distributions theory (Clark, 1973;

Tauchen and Pitts, 1983).² This theory states that both trading volume and the return on a specific asset are governed by a latent mixing process, which is interpreted as the rate of information arrival. A common hypothesis is that volumes should have some information content on time-varying volatility. A related strand of research is based on the analysis of microstructures of financial markets (Admati and Pfleiderer, 1988). The recent availability of very high frequency data of financial markets allows researchers to test these theoretical models (Bollerslev and Domowitz, 1993; Brock and Kleidon, 1992).

The structure of this chapter is as follows. In Section 2.1 we will present and discuss some models from the ARCH family, in particular the EGARCH model that was developed by Nelson (1991). In Section 2.2 we will discuss the Stochastic AutoRegressive Volatility (SARV) class of models that has been introduced by Andersen (1994b). SARV models are very flexible with respect to modelling volatility. Conditional heteroskedastic models are only a small part of the possible models that can be put in this class. In Section 2.3 we will discuss volatility models that make full use of the stochastic structure that the SARV class allows. Special focus will be on the lognormal Stochastic Volatility model, which is the most widely studied SV model in the literature. On the basis of implied statistical properties of both ARCH and SV models we make a comparison in Section 2.4. Section 2.5 concludes this chapter.

2.1 Conditional Heteroskedasticity Models

Recently, some very good survey articles on ARCH models have appeared in the literature (Bollerslev, Chou and Kroner, 1992; Bollerslev, Engle and Nelson, 1993, and Pagan, 1995). In this section we will not attempt to redo the work that these authors have done. Instead, we will highlight some aspects of ARCH models that will be important for the discussion later in this chapter.

The main motivation for developing conditionally heteroskedastic models by Engle (1982) was that existing linear time series models were considered to be inappropriate for obtaining "good" forecast intervals. "Good" means that both the conditional mean as well as the conditional variance of a time series should play an important role in making forecasts. Although originally developed in a macroeconomic setting, ARCH models have appeared to be extremely valuable in finance. Most financial time series alternate periods of relative calm with periods of increased volatility. Engle's ARCH(p) model has the following form:

$$\epsilon_t | I_{t-1} \sim ID(0, \sigma_t^2) \quad (2.1)$$

²See also Chapter 5.

$$\sigma_t^2 = \alpha_0 + \sum_{j=1}^p \alpha_j \epsilon_{t-j}^2, \quad (2.2)$$

where I_{t-1} is an information set that contains all information through time $t-1$. In the first equation σ_t^2 is the conditional variance. Define $\epsilon_t \equiv y_t - \mu_t$, where μ_t is the conditional mean of the data generating process $\{y_t\}_{t=1}^T$. The conditional mean μ_t is measurable with respect to I_{t-1} . A sufficient condition to ensure positive conditional volatility is to let the coefficients α_j obey the restrictions $\alpha_j \geq 0$ ($j = 1, \dots, p$) and $\alpha_0 > 0$. Nelson and Cao (1992) show that these restrictions can be weakened. Covariance-stationarity obtains if and only if $\sum_{j=1}^p \alpha_j < 1$. In many empirical applications the standardized error terms $u_t \equiv \epsilon_t / \sigma_t$ are assumed to be drawn from a normal, Student- t or generalized error (Nelson, 1991) probability distribution function. Note that the ϵ_t are dependent through time. This property of ARCH models explains their success in the modelling of financial time series.

Under conditional normality the ARCH model shows that the temporal behavior of $y_t = \mu_t + \sigma_t u_t$ is determined by conditional first and second order moments only. Define $\hat{\mu}_t$ and $\hat{\sigma}_t$ to be the estimated conditional mean and standard deviation, respectively. Then $\hat{u}_t \equiv (y_t - \hat{\mu}_t) / \hat{\sigma}_t$ can be interpreted as a residual that has a probability density function which is invariant if the volatility model is given by $y_t = \mu_t + \sigma_t u_t$ (Andersen, 1994b). In the ARCH model both conditional processes are observable with respect to I_{t-1} . This is an important feature of any model in the conditional heteroskedastic volatility class.

The Generalized ARCH (GARCH) model was introduced by Bollerslev (1986). The main reason for the extension of Engle's (1982) ARCH model was that in empirical applications it turned out that a long lag structure (large p) was needed. The conditional volatility equation of a GARCH(p, q) model is given by

$$\sigma_t^2 = \alpha_0 + \sum_{j=1}^p \alpha_j \epsilon_{t-j}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2. \quad (2.3)$$

Positive conditional volatility is guaranteed if $\alpha_j, \beta_j \geq 0$ and $\alpha_0 > 0$. The process ϵ_t^2 is covariance-stationary if and only if $\sum_{j=1}^p \alpha_j + \sum_{j=1}^q \beta_j < 1$. Again, results in Nelson and Cao (1992) show that these restrictions can be weakened. Engle and Bollerslev (1986) were the first authors to apply the GARCH model to financial time series. It turned out that the GARCH(1,1) model captures the temporal dependence in conditional variance quite well. In many studies this finding has been confirmed (Bollerslev, Chou and Kroner, 1992).

In both ARCH and GARCH models restrictions have to be placed on the parameters to keep the conditional volatility positive. This could create problems in estimating these models. To ensure positivity we could model the conditional volatility as $\ln \sigma_t^2$ instead of σ_t^2 . This is the idea of Nelson's (1991) Exponential GARCH (EGARCH). The full

specification that he chooses to work with is the EGARCH(p, q) model

$$\ln \sigma_t^2 = \alpha_0 + \sum_{j=1}^p \beta_j \ln \sigma_{t-j}^2 + \sum_{j=1}^q \gamma_j (\theta_0 u_{t-j} + \gamma_0 [|u_{t-j}| - E|u_t|]) \quad (2.4)$$

The last part of this equation captures the so-called leverage phenomenon. This is an asymmetric relation between the return on an asset and volatility changes. The sign of the leverage effect depends on the sign of θ_0 . Shocks in u_t have different effects on volatility: if there is negative news volatility tends to rise whereas positive news lowers volatility (Black, 1976).³

2.2 A Unified Framework of Volatility

A useful guide for classification of volatility models is the article of Andersen (1992), which explores the measurability properties of volatility with respect to certain information sets. Based on these ideas Andersen (1994b) introduces the class of Stochastic Autoregressive Volatility (SARV) models. This model class incorporates both classic ARCH models, stochastic extensions to the familiar conditional heteroskedastic models and more general SV models. Consequently, SARV models are a natural device to study the relationships between models from the autoregressive heteroskedasticity class and SV models.

For ease of exposition we will restate the definition of polynomial SARV models of Andersen (1994b). A univariate process y_t is a polynomial SARV model if it satisfies the following conditions:

- (i) $y_t = \mu_t + \sigma_t u_t$,
- (ii) $\sigma_t^k = g(Z_t)$, $k = 1, 2$ with g an increasing function,
- (iii) $Z_t = h(Z_{t-1}, w_t) = a_0 + a_1 Z_{t-1} + [a_2 + a_3 Z_{t-1}] w_t$,
- (iv) $g(Z_t) = \sum_{j=0}^{\infty} b_j Z_t^j$,
- (v) w_t is i.i.d. $(1, \sigma_w^2)$, $w_t \in (-a_1/a_3, \infty)$,
- (vi) u_t is independent of w_{t-j} , $j \geq 0$.

The volatility process y_t in (i) is a very general process that is governed by conditional first and second moments, which is a typical requirement for any volatility model. It is assumed

³See also Chapter 5.

that μ_t is measurable with respect to information up to time $t-1$. The process $\{Z_t\}$ is called a SARV process and is a latent process that drives the volatility. Condition (ii) implies that volatility dynamics can be modelled for σ_t and σ_t^2 alike. From (iii) we see that the SARV process is restricted to a first order dynamic structure. Higher order dynamics can be readily introduced at the cost of losing parsimony. The following restrictions apply to the parameters in (iii): $a_1, a_2, a_3 \geq 0, a_1 + a_3 > 0, a_2 + a_3 > 0$. If $a_0 = a_2 = 0$, then $Z_t = 0$ for all t . It is possible to restrict the infinite sum in (iv) by setting $b_j = 0$ for $j \geq J$, with J a positive integer. The shocks w_t in (v) are normalized to have unit mean and are restricted to be drawn from the interval $(-a_1/a_3, \infty)$ to ensure that $a_1 + a_3 w_t > 0$. This latter condition is necessary and sufficient to prove stationarity and ergodicity of the SARV model (Andersen, 1994).

The conditional heteroskedastic models from the previous section can be fit in the polynomial SARV model class quite easily. First consider the GARCH(1,1) model. Higher order GARCH models do not fit in the SARV framework with a first-order latent process Z_t (see condition (iii)). If we set $g(Z_t) = Z_t$, $k = 2$, $a_0 = \alpha_0$, $a_1 = \alpha_1$, $a_2 = 0$, $a_3 = \beta_1$, $w_t = u_{t-1}^2$, substitution in the SARV definition leads to the desired result. Note that the random variables w_t are replaced by u_{t-1} which is measurable with respect to the information set I_{t-1} . The SARV representation of ARCH(1) follows analogously when we set $a_1 = 0$ in the GARCH representation. A similar exercise can be done for the EGARCH(1,1) model. Here we have to set $g(Z_t) = \exp(Z_t)$, $k = 2$, $a_0 = \alpha_0$, $a_1 = \beta_1$, $a_2 = \gamma_1$, $a_3 = 0$ and $w_t = \theta_0 u_{t-j} + \gamma_0 [|u_{t-j}| - E|u_t|]$.

Andersen shows that the polynomial SARV process $\{Z_t\}$ has a set of desirable statistical properties that makes it suitable for standard (method of moments) estimation techniques. He shows that SARV processes are strictly stationary and ergodic. Also, under some mild restrictions on the parameters, unconditional moments exist.

2.3 Stochastic Volatility Models

Stochastic volatility models have their origin in theoretical finance. In the continuous time literature on topics like portfolio diversification (Merton, 1990) and in particular the pricing of derivative instruments (Hull and White, 1987; Johnson and Shanno, 1987; Chesney and Scott, 1989; Melino and Turnbull, 1990 among others) it has been understood that constant volatility, as in the classic Black and Scholes (1973) model, is too restrictive for models to have sufficient empirical support.

Several models for continuous time return processes have been proposed in the literature. Here we will present a typical specification, that has all the important features of a return process with stochastic volatility. Let the spot asset price S_t follow a diffusion process

$$dS_t = \alpha S_t dt + \sigma_t S_t dw_{1t}, \quad \alpha > 0,$$

with σ_t as a measure for volatility, and w_{1t} a Brownian motion. By Ito's lemma we find that

$$d \ln S_t = \left(\alpha - \frac{1}{2} \sigma_t^2 \right) dt + \sigma_t dw_{1t} \quad (2.5)$$

This equation can be interpreted as a process for the instantaneous return on the asset S . The volatility can be allowed to change over time according to the Ornstein-Uhlenbeck (O-U) process

$$d\sigma_t^2 = -\beta (\sigma_t^2 - \mu) dt + \gamma \sigma_t^2 dw_{2t}, \quad \beta, \gamma > 0,$$

which can again be transformed by Ito's lemma to

$$d \ln \sigma_t^2 = \left(\frac{\beta \mu}{\sigma_t^2} - \beta - \frac{1}{2} \gamma^2 \right) dt + \gamma dw_{2t} \quad (2.6)$$

The Brownian motion dw_{2t} is correlated with dw_{1t} by a constant correlation ρdt . The parameter μ is the long-term mean of the volatility process and β governs the adjustment to the mean. The O-U process has the desirable property that it is always positive, because the upward drift is sufficiently large with respect to the stochastic part $\gamma \sigma_t dw_{2t}$ (Shimko, 1992). Other specifications for the volatility process have been given in the literature. It should be remembered though, that most continuous time model specifications in the literature are chosen in order to arrive at closed-form solutions of the partial differential equations that stand at the basis of derivative security pricing.

In order to test the specification of the continuous time model and to estimate the parameters we need to confront the model with data that are sampled in discrete time. We therefore need discrete time approximations of equations (2.5) and (2.6). The simplest approximation can be obtained by applying so-called Euler approximations.⁴ Let us first take a look at (2.5). Suppose that we have $N+1$ observations that are equally spaced on the time interval $[t_0, T]$. Let δ be the step size between two successive observations, *i.e.* $\delta = (T - t_0)/N$. Define τ_n to be the time at which an observation is made: $\tau_n = t_0 + n\delta$ ($n = 0, \dots, N$). The Euler approximation then becomes

$$\ln S_{n+1} - \ln S_n = \left(\alpha - \frac{1}{2} \sigma_{\tau_n}^2 \right) \delta + \sigma_{\tau_n} (W_{1,\tau_{n+1}} - W_{1,\tau_n}), \quad n = 0, \dots, N-1, \quad (2.7)$$

⁴Nelson and Foster (1994) analyze conditions under which discrete time approximations are optimal. Drost, Nijman and Werker (1994) investigate tests for the hypothesis that a particular discrete time series has been generated by a diffusion process.

where S_{n+1} is the value of the time series at time τ_{n+1} . The initial value of the process is given by $S_0 = S_{t_0}$. The random increments $\Delta W_{1n} \equiv W_{1,\tau_{n+1}} - W_{1,\tau_n}$ are independent Gaussian random variables with mean zero and variance equal to δ . From (2.7) we see that the return on the asset depends on volatility in a dual way. Past volatility affects both the mean and the variance of the return process.⁵

The same discretization can be performed on the volatility process. The equation becomes

$$\ln \sigma_{n+1}^2 - \ln \sigma_n^2 = \left(\frac{\beta\mu}{\sigma_n^{2(a)}} - \beta - \frac{1}{2}\gamma^2 \right) \delta + \gamma \Delta W_{2n}, \quad n = 0, \dots, N-1, \quad (2.8)$$

where $\Delta W_{2n} \equiv W_{2,\tau_{n+1}} - W_{2,\tau_n}$. The correlation between ΔW_{1n} and ΔW_{2n} is equal to $\rho\delta$. The increments in logarithmic volatility depend on the level of volatility: if volatility is relatively high, volatility growth will be slowed, all else equal. This shows that the mean-reverting nature of volatility is kept if we make the switch from a continuous time to a discrete time model.

Many applications of the SV model in the literature have a return process in discrete time that ignores the volatility dependence of the conditional mean. If we redefine the time index t as a time indication in discrete time a typical model would be

$$\ln S_{t+1} - \ln S_t = \alpha^* + \sigma_t \epsilon_{t+1}, \quad (2.9)$$

where the ϵ_{t+1} is an identically and independently distributed normal random variable with mean zero and variance equal to one. A common specification of the volatility process is an AR(1) process:

$$\ln \sigma_{t+1}^2 = \mu^* + \phi (\ln \sigma_t^2 - \mu^*) + \eta_{t+1}. \quad (2.10)$$

The parameter ϕ is required to be smaller than one in order to keep the volatility mean-reverting. This volatility model is also called the lognormal model. The error term η_{t+1} is usually taken to be identically and independently normally distributed with mean zero and variance σ_η^2 . The correlation between ϵ_{t+1} and η_{t+1} is equal to ρ^* .

In Equation (2.9) we see that the volatility at time t acts as a scale factor for the idiosyncratic shock to the return. This is a direct result from the Euler discretization. Remember that the equations (2.9) and (2.10) are statistical models but with the desirable property that they converge to continuous time models used in theoretical finance.⁶ For practical purposes, the SV model, as given above, does not differ much from conditional heteroskedastic models. Although the volatility equation (2.10) is governed by a stochastic process, it is measurable with respect to the information set at time t . When we compare (2.9) with condition (i) of a SARV process (with $y_{t+1} \equiv \ln S_{t+1} - \ln S_t$) we see that

⁵This specification is reminiscent of the GARCH-in-mean model of Engle, Lilien and Robins (1987).

⁶Nelson (1990) gives conditions under which these models converge to continuous time limits.

straightforward application of the Euler approximations does not give us a pure stochastic volatility model. Instead, replace σ_t in (2.9) with σ_{t+1} . It can be shown that this volatility representation will also converge to the continuous time volatility process (2.5), when we let δ go to zero.

The logarithmic transformation of the volatility forces the volatility process to be positive. The reasons for the predominance of the AR(1) model for volatility might be that efficient estimation of even the simplest SV models is a real challenge for econometricians. We will discuss the estimation of SV models in the next chapter. Here we will concentrate on some simple properties of the SV model, that allow us to compare it with the conditional heteroskedastic models from the previous section. To this end we use the SARV class of models again.

First, make the transformation $h_t \equiv \ln \sigma_t$. Consider the following SV model

$$\begin{aligned} y_{t+1} &= \alpha^* + \exp\left(\frac{1}{2}h_{t+1}\right)u_{t+1}, \\ h_{t+1} &= \mu^* + \phi(h_t - \mu^*) + \eta_{t+1}. \end{aligned} \quad (2.11)$$

To put the lognormal SV model in the SARV specification we first consider the volatility equation in (2.11). Rewrite this equation as

$$h_{t+1} = \mu^* + \phi(h_t - \mu^*) + w_{t+1} - 1, \quad (2.12)$$

where w_{t+1} has unit mean, as in condition (v) of the SARV definition. Then

$$\begin{aligned} \epsilon_{t+1} &\equiv y_{t+1} - \alpha^* \\ &= \exp\left(\frac{1}{2}h_{t+1}\right)u_{t+1} \\ &= \exp\left(\frac{1}{2}\mu^* + \frac{1}{2}\phi(h_t - \mu^*) - \frac{1}{2}\right)\left(\exp\left(\frac{1}{2}w_{t+1}\right)u_{t+1}\right) \\ &\equiv \bar{\sigma}_t \bar{u}_{t+1}, \end{aligned}$$

where $\bar{\sigma}_t \equiv \exp(Z_t)$ is the SARV function. This is again a volatility representation. The term $\exp(Z_t)$ conforms with condition (iv) since we can write $\exp(Z_t) = \sum_{j=0}^{\infty} Z_t^j / j!$. The variable \bar{u}_{t+1} is a random variable that is made up from a mixture of the two sources of randomness in the original SV model. The fact that $\bar{u}_{t+1} \equiv \exp(w_{t+1}/2)u_{t+1}$ is random is the main difference with the SARV representation of ARCH models.

2.4 Moment Comparison

In this section we will provide a further comparison of ARCH and SV models based on unconditional non-central moments. Pagan (1995) shows that the autocorrelation function

of the squared returns implied by the lognormal SV model gives a closer fit to the empirical autocorrelations than a GARCH(1,1) model. This finding warrants a closer investigation of SV models and a further comparison between ARCH and SV models.

We choose the three most widely used volatility models that have been applied to financial time series: GARCH(1,1), EGARCH(1,1) and the lognormal SV model with AR(1) volatility. To facilitate the discussion we repeat the exact specifications of the models. We assume that the conditional mean μ_t is zero. The observation equation that we use for all three models is

$$y_t = \sqrt{h_t} u_t, \quad u_t \sim N(0, 1)$$

The volatility equations for the three models are

$$\begin{aligned} \text{GARCH : } h_t &= \alpha_0 + \alpha_1 y_{t-1}^2 + \alpha_2 h_{t-1}, \\ \text{EGARCH : } \ln h_t &= \beta_0 + \beta_1 \ln h_{t-1} + \beta_2 u_{t-1}, \\ \text{SV : } \ln h_t &= \gamma_0 + \gamma_1 \ln h_{t-1} + \eta_t, \quad \eta_t \sim N(0, \sigma_\eta^2). \end{aligned}$$

The EGARCH(1,1) volatility equation has no leverage terms. We require that $|\beta_1| < 1$. The lognormal SV model has the restriction that $|\gamma_1| < 1$. We assume that u_t and η_t are jointly independently distributed. Note that the number of unknown parameters is equal to three in all models. We will compute the first 4 non-central unconditional moments. Since the u_t are independent of all other random variables in the models, we have $E[y_t^m] = 0$ for all m odd.

In case of computation of moments of the EGARCH and SV model it will be useful to know the unconditional distribution of $\ln h_t$. We find

$$\begin{aligned} \text{EGARCH : } \ln h_t &\sim N(\beta_0/(1 - \beta_1), \beta_2^2/(1 - \beta_1^2)) \equiv N(\mu_\beta, \sigma_\beta^2), \\ \text{SV : } \ln h_t &\sim N(\gamma_0/(1 - \gamma_1), \sigma_\eta^2/(1 - \gamma_1^2)) \equiv N(\mu_\gamma, \sigma_\gamma^2). \end{aligned}$$

From this it follows that h_t is lognormally distributed. Note that the EGARCH and the SV model have similar unconditional distributions of the volatility. However, in a conditional analysis the two models will differ as a consequence of the measurability of the stochastic terms of the model. For example, the predictive density of the EGARCH model will contain the realized error \hat{u}_{t-1} , whereas in the SV model η_t is not in the information set.

The second moments are given by $Ey_t^2 = Eh_t u_t^2 = Eh_t Eu_t^2 = Eh_t$. This leads to

$$\begin{aligned} \text{GARCH : } Eh_t &= \alpha_0/(1 - \alpha_1 - \alpha_2) \\ \text{EGARCH : } Eh_t &= \exp\left(\mu_\beta + \frac{1}{2}\sigma_\beta^2\right) \\ \text{SV : } Eh_t &= \exp\left(\mu_\gamma + \frac{1}{2}\sigma_\gamma^2\right) \end{aligned}$$

The restriction that $\alpha_1 + \alpha_2 < 1$ in the GARCH model is necessary for the second moment to exist.

The fourth moments can be constructed from $Ey_t^4 = Eh_t^2 u_t^4 = Eh_t^2 E u_t^4 = 3Eh_t^2$. Then

$$\text{GARCH : } 3Eh_t^2 = 3[\alpha_0 + 2\alpha_0(\alpha_1 + \alpha_2)/(1 - \alpha_1 - \alpha_2)] / (1 - 3\alpha_1^2 + \alpha_2^2 + 2\alpha_1\alpha_2)$$

$$\text{EGARCH : } 3Eh_t^2 = 3 \exp(2\mu_\beta + 2\sigma_\beta^2)$$

$$\text{SV : } 3Eh_t^2 = 3 \exp(2\mu_\gamma + 2\sigma_\gamma^2)$$

The moments for EGARCH and SV clearly indicate that these models imply excess kurtosis with respect to the normal distribution, which has a fourth moment that is equal to 3. The GARCH model also has excess kurtosis if $3\alpha_1^2 + \alpha_2^2 + 2\alpha_1\alpha_2 < 1$.

2.5 Conclusions

In this chapter we have introduced the two most widely used volatility models in finance. We have seen that the data generating process in any volatility model is governed by the first two moments. The information set on which the moments are minimally measurable (Andersen, 1992) determines what type of volatility process we are dealing with. The class of SARV models permits a clear distinction between conditional heteroskedastic models and stochastic volatility models. The SARV representation also shows that many more volatility models can be developed. In particular, direct stochastic extensions to the familiar ARCH class can readily be specified.

Chapter 3

The Estimation of Stochastic Volatility Models

Many financial time series show time-varying volatility. The most popular way to describe the time variation has been the AutoRegressive Conditional Heteroskedasticity (ARCH) model introduced by Engle (1982), or one of its many variants (GARCH, Bollerslev, 1986; EGARCH, Nelson, 1991). An alternative approach, which has become more popular recently, is the Stochastic Volatility (SV) model, where the variance is modeled as an unobservable component that follows some stochastic process. In this chapter we will take up the task of estimating SV models. Although this has turned out to be a challenging task for econometricians, several techniques have already been developed. The statistical research of SV models still is in a state of flux. Undoubtedly, new techniques will emerge in the nearby future. In this chapter we will discuss the techniques that have been used successfully, and we will present an extension to one of the estimation methods.

From an econometric viewpoint the practical drawback of the SV model is the intractability of the likelihood function. Because the variance is an unobserved component, the likelihood is only available in the form of a multiple integral. Moreover, it turns out that Quasi Maximum Likelihood and Method of Moments estimators are not efficient, see Jacquier, Polson and Rossi (1994) and Andersen and Sørensen (1994).

One purpose of this chapter is to compare different estimation techniques applied to an empirical dataset consisting of weekly exchange rate changes for four currencies over a 20 year period. The comparison is made in terms of both parameter and volatility estimates. The estimation techniques for SV models differ in their distributional assumptions about exchange rate innovations. Most of the computational problems stem from the assumption that the innovation in the (nonlinear) measurement equation has a normal distribution,

which translates to an awkward log chi-square distribution when the model is written in a linear state-space form. This implication is ignored in the Quasi Maximum Likelihood method, but fully implemented in the methods of Jacquier, Polson and Rossi (1994) and Kim and Shephard (1994).

The normality assumption on the measurement error is not only computationally impracticable, it is probably also empirically unjustified. The unconditional distribution of exchange rate news is fat-tailed as has been widely established in the literature. Like ARCH models, stochastic volatility can explain part of the fat-tailedness (Pagan, 1995). But given the evidence for ARCH models one would expect that time varying volatility does not fully account for the tail behavior, in line with Baillie and Bollerslev (1989) and Engle and Bollerslev (1986). The second purpose of this chapter is to replace the normally distributed innovation with a specification that is computationally more tractable and empirically viable.

Estimation of SV models consists of two parts: parameter estimation, and estimation of the latent volatility time series. Methods that work well for estimating the parameter vector, are not necessarily also good in estimating volatility series. For finance applications the main interest is in the volatility time series itself. The series is estimated by some smoothing algorithm, which also produces standard errors of the volatility estimate. This enables us to compare the different models and estimation techniques with respect to the estimated volatility series. Focussing directly on the output of the model – the volatility series – sheds light on issues like the efficiency gain from a simulation smoother over the Gaussian linear Kalman smoother, and the effect of some forms of misspecification. Because the interest is in estimating the latent volatility series at every time period, asymptotic arguments are of limited value. Consequently, distributional assumptions become important for this purpose.

Like the GARCH(1,1) model in the conditional heteroskedasticity literature, most work on stochastic volatility deals exclusively with the basic univariate first order autoregressive volatility model. For stock prices Gallant, Hsieh and Tauchen (1994) conclude that this basic model is severely misspecified, not only because of distributional assumptions on innovations, but also because of its dynamic specification. Similar results might hold for exchange rates. The third aim of this chapter is to perform a battery of diagnostic Lagrange Multiplier (LM) tests to search for deviations from the basic SV model.

The remainder of the chapter is organized as follows. A discussion of the estimation techniques that have been developed in the literature and the corresponding technical issues follows in Section 3.1. A broad classification of estimation methods is given by simulation

estimators on the one side and non-simulation estimators on the other. We present the model specification of the mixture model and the estimators we use in Section 3.2. We consider four different estimators in our empirical analysis: Quasi Maximum Likelihood, two different simulation techniques based on the EM algorithm, and a Bayesian method. In Section 3.3 we present the parameter estimation results and diagnostic test results. Section 3.4 reports the results for the estimation of the latent volatility series. Section 3.5 contains some conclusions.

3.1 Estimation Techniques

In order to understand the basic statistical problems that SV models pose consider a very simple nonlinear state-space model with an AR(1) latent volatility specification

$$s_t = \exp \left[\frac{1}{2}(\beta + h_t) \right] \epsilon_t, \quad t = 1, \dots, T, \quad (3.1)$$

$$h_{t+1} = \phi h_t + \eta_t, \quad (3.2)$$

where s_t is the return of a financial asset and h_t is the conditional logarithmic volatility up to a constant. The error terms ϵ_t and η_t are assumed to be uncorrelated and identically distributed with means zero and variances 1 and σ_η^2 , respectively. The constant term β can be interpreted as a benchmark volatility level. If $\phi < 1$ then β can be interpreted as the unconditional volatility, because in that case the volatility process is strictly stationary. Consequently, s_t is also stationary, since the product of two strictly stationary series is again strictly stationary. The hyperparameters of this model are contained in the vector $\theta \equiv (\beta, \phi, \sigma_\eta^2)'$. Let $S_T = (s_1, \dots, s_T)'$ and $H_T = (h_1, \dots, h_T)'$. The conditional density of the observations S_T given the latent variables H_T and the parameters is defined to be $f(S_T|H_T; \theta)$. The conditional density of all the state variables given the parameters is $f(H_T|\theta)$. Then, the likelihood function of the process (3.1) and (3.2) is

$$L(S_T; \theta) \propto \int f(S_T|H_T; \theta) f(H_T|\theta) dH_T. \quad (3.3)$$

The estimation of parameters based on the likelihood of a SV model amounts to integrating out the latent variables H_T . The typical length of financial time series prohibits direct evaluation of the likelihood. Therefore, techniques have been developed to try to deal with the integration problem in a tractable way. Comparing the likelihood with the objective function in standard Bayesian analysis leads to the observation that $f(H_T|\theta)$ can be interpreted as a prior over the latent variables H_T . In other words, the specification of the SV model itself is a prior for H_T (Jacquier, Polson and Rossi, 1994).

As we noted in Chapter 2 SV models are closely related to continuous time models which

have a firm base in financial asset pricing theory. Until recently, direct estimation of continuous time models in finance has been given little attention though. Notable exceptions are Engle and Lee (1994), who base their estimation on the indirect inference approach of Gourieroux, Monfort and Renault (1993). Ghysels, Gourieroux and Jasiak (1995) and Conley, Hansen, Luttmer and Scheinkman (1994) study subordinated processes in continuous time. Drost, Nijman and Werker (1994) analyze tests for the hypothesis that a series is generated by a continuous time model. A clear reason for the apparent lack of studies on direct continuous time estimation cannot be found easily. The fact that observations on most economic and financial time series are sampled at relatively low (at most daily) frequency, together with the fact that existing time series models in discrete time perform satisfactorily enough with respect to in- and out-of-sample forecasting, may have led many researchers to discard continuous time estimation from the outset. Also the technical requirements might be prohibitive in applying this kind of estimation. The advent of data with very high frequencies might shift the interest towards continuous time estimation methods in finance in the future. In this chapter we will exclusively deal with estimation methods in discrete time.

This section is divided into two parts. In the first part (Section 3.1.1) we will discuss the Quasi Maximum Likelihood class of estimators and estimators based on the Method of Moments. In principal, these methods do not rely on simulation exercises, although some modifications may. The second part of this section (Section 3.1.2) will discuss estimation models that are primarily based on simulation methods, in particular Monte Carlo Markov Chain (MCMC) samplers. Our empirical work in the subsequent sections has been skewed towards the simulation type estimators for SV models.

•

3.1.1 Non-Simulation Estimators

In this section we will look at two estimation methods that approach the integration problem (3.3) in such a way that simulation is not necessary. The first method approximates the state-space model (3.1) and (3.2) by a linear Gaussian state-space. Consequently, standard Kalman filter and smoothing algorithms can be employed to find parameters and volatility series. This method is called Quasi Maximum Likelihood estimation. The second method is based on the Method of Moments estimation procedures. It will turn out that these methods do not always deliver optimal inference. Especially, parameter estimates turn out to be statistically inefficient with respect to the simulation based estimators presented in the next section.

Quasi Maximum Likelihood Estimation

The Quasi Maximum Likelihood (QML) method for estimating SV models is perhaps the single most used. It was first introduced by Nelson (1988) and further explored in the article of Harvey, Ruiz and Shephard (1994). Its attraction has at least three reasons. First, it is relatively easy to apply, since only well-known and established estimation techniques are used. Second, it is fast compared to other methods. Third, it serves as a starting point for more elaborate estimation techniques, like the simulation methods of the next section. However, QML imposes some severe restrictions on SV models. In this subsection we will look at the statistical implications of the QML method.

The first step in the QML method is to linearize the measurement equation (3.1). Squaring all elements from the observation vector S_T and taking logarithms, we obtain

$$\ln s_t^2 = \beta + h_t + \ln \epsilon_t^2, \quad t = 1, \dots, T. \quad (3.4)$$

Note that the system (3.4) and (3.2) now comprises a linear albeit non-Gaussian state-space. Harvey, Ruiz and Shephard (1994) (HRS henceforth) assume that $\xi_t \equiv \ln \epsilon_t^2$ is normally distributed, with mean and variance equal to $E[\xi_t] = -1.2704$ and $\text{Var}[\xi_t^2] = \frac{1}{2}\pi^2$, respectively, under the added assumption that ϵ_t is normal. For ϵ_t and η_t bivariate normal with some unknown correlation, HRS show that the transformed error term ξ_t is always uncorrelated with η_t . This is due to the fact that the joint distribution of ϵ_t and η_t is symmetric.

HRS suggest to ignore the non-normality and estimate the system by QML. The steady-state Kalman filter recursions then provide an expression for the conditional (log)-variance of s_t ,

$$\hat{h}_{t+1|t} = \beta(1 - \phi) + \frac{2\phi\bar{P}^2}{2\bar{P}^2 + \pi^2} \ln s_t^2 + \frac{2\phi\pi^2}{2\bar{P}^2 + \pi^2} \hat{h}_{t|t-1},$$

where \bar{P}^2 is the solution to the Riccati equation

$$\bar{P}^2 = \phi^2 \left(\frac{2}{\pi^2} + \frac{1}{\bar{P}^2} \right) + \sigma_\eta^2.$$

Using the properties of the log-normal distribution the conditional variance of s_{t+1} (given information up to time t) is given by

$$\text{Var}_t(s_{t+1}) = \exp \left(\frac{1}{2} \hat{h}_{t+1|t} + \frac{1}{8} \bar{P}^2 + \frac{1}{2} \pi^2 \right),$$

which establishes the relation between the SV process and an equivalent exponential ARCH process. Estimation of the two processes differs, however. The QML of the SV model involves maximization of the objective function

$$F_1 = -\frac{T}{2} \ln \psi^2 - \frac{1}{2\psi^2} \sum_{t=1}^T \left(\ln s_t^2 - \hat{h}_{t|t-1} - \beta \right)^2,$$

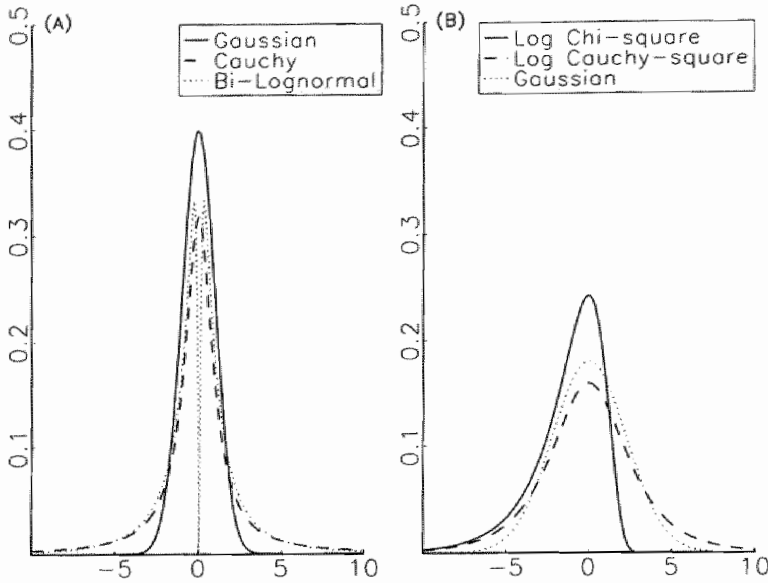


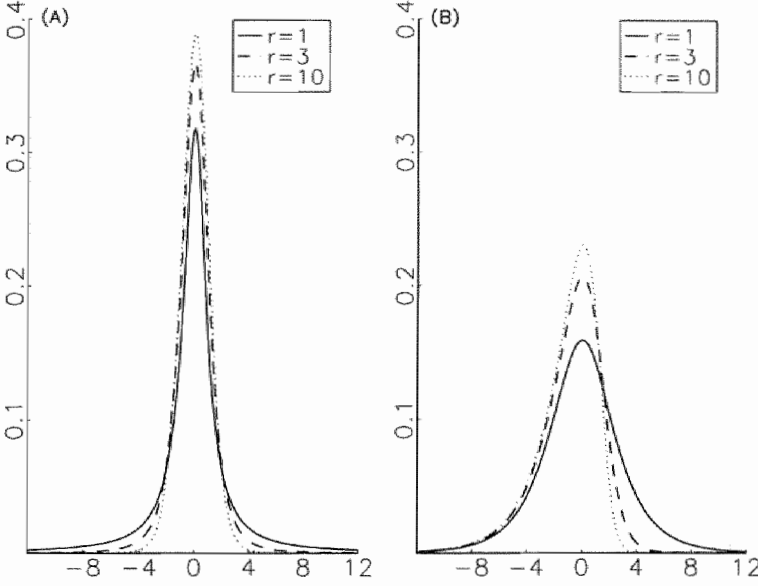
FIGURE 3.1: Probability Density Functions of ϵ_t (A) and $\xi_t = \ln \epsilon_t^2$ (B).

where $\psi^2 \equiv \bar{P}^2 + \frac{1}{2}\pi^2$ is the steady state innovation variance of $\ln s_t^2$. In contrast, an ARCH model would be estimated by maximizing the objective function

$$F_2 = -\frac{1}{2} \sum_{t=1}^T \left(\ln \text{Var}_t(s_{t+1}) + \frac{s_t^2}{\text{Var}_t(s_{t+1})} \right).$$

The difference between the two criterion functions stems from the assumption on which innovation is taken to be Gaussian. With ARCH, s_t is assumed conditionally normal with zero mean and time-varying conditional variance, while for the QML of the SV model the logarithm of the squared exchange rate is assumed Gaussian with time varying mean and constant variance. Otherwise the two models are equivalent.

The distribution of ϵ_t plays an important role in the model, since it determines the mean and variance of $\ln s_t^2$. In the case of assuming a normal distribution for ϵ_t we obtain a very skewed distribution of ξ_t , see Figure 3.1. But the skewness of ξ_t is an empirical matter, just like the normality of ϵ_t . In the empirical ARCH literature Engle and Bollerslev (1986) find that the standardized exchange rate innovations are still leptokurtic, which leads them, and Baillie and Bollerslev (1991), to consider the Student- t distribution as an alternative. If ϵ_t is fat-tailed the implied distribution of $\ln \epsilon_t^2$ becomes less skewed than the log chi-squared. In the extreme case that ϵ_t has a Cauchy distribution, the implied density of ξ_t

FIGURE 3.2: Student- t Densities of ϵ_t (A) and $\xi_t = \ln \epsilon_t^2$ (B).

is the symmetric function

$$p(\xi) = \frac{\exp\left(\frac{1}{2}\xi\right)}{2\pi(1 + \exp(\xi))}.$$

In this case the expectation of ξ_t is equal to zero while $\text{Var}(\xi_t) = \pi^2$. The distribution $p(\xi)$ is plotted in Figure 3.1 together with the normal and the log chi-squared. Given the symmetry and exponential tails of $p(\xi)$, we would expect that the QML performs much better in this case. Figure 3.1 also shows the case of ξ_t being normally distributed, implying that ϵ_t has a bi-lognormal distribution

$$p(\epsilon) = \frac{1}{|\epsilon|\pi\sqrt{\pi}} \exp\left(-\frac{(\ln \epsilon^2)^2}{\pi^2}\right).$$

Normality of ξ_t implies that very small innovations in ϵ_t are relatively unlikely, see the dip in the plotted density. If we assume that ϵ_t is distributed as a Student- t with r degrees of freedom the log Student- t squared distribution is

$$p(\xi) = \frac{\Gamma\left(\frac{1}{2}(r+1)\right)}{\Gamma\left(\frac{1}{2}r\right)} \frac{\sqrt{\exp(\xi)}}{\left(1 + \frac{\exp(\xi)}{r}\right)^{\frac{1}{2}(r+1)}}$$

In Figure 3.2 we plotted the Student- t densities. It can be seen that the lower the degrees of freedom parameter r , the more symmetric the distribution function of ξ_t becomes. Note that the Student- t distribution approaches the Cauchy distribution if we let r go to zero.

Finally, note that the constant term β in (3.4) can be redefined to include the mean of ξ_t . From an estimation point of view these two constants are not separately identifiable as free parameters. A possible way to identify $E(\xi_t)$ is through the estimated variance of ξ_t . For example, if $\text{Var}(\xi_t) = \frac{1}{2}\pi^2$, ξ_t might be generated by the log chi-squared, in which case $E(\xi_t) = -1.2704$. Alternatively, if $\text{Var}(\xi_t) = \pi^2$, and the residuals ξ_t are symmetric, we might be dealing with the Cauchy distribution for ϵ_t , so that $E(\xi_t) = 0$. In this case the conditional variances of exchange rates do not exist, but there is still a meaningful way to describe time-varying volatility through the h_t process, *i.e.* $\exp\left(\frac{1}{2}h_t\right)$ still functions as a scale parameter.

Method of Moments Estimation

Besides the QML method for estimating SV models, the method of moments (MM) has been extensively applied to SV models. In words, the MM objective function minimizes a weighted squared sum of the differences between sample and implied moments. Taylor (1986) and Melino and Turnbull (1990) were the first to apply MM to SV models. Other articles include Andersen and Sørensen (1994), who perform an extensive Monte Carlo analysis and Andersen (1994b), who applies MM to the SARV class, that we discussed in Chapter 2. Gallant, Hsieh and Tauchen (1994) perform estimation of SV models using a version of MM that is more efficient.

The first step in every MM estimation exercise is the choice and calculation of implied moments. In some cases moments cannot be calculated analytically and one has to resort to simulating the relevant moments (Duffie and Singleton, 1993). In order to get consistent parameter estimates, the number of moments needed to estimate the parameter vector θ , is equal to the number of parameters in θ . However, Hansen (1982) has shown that the number of moments may be increased relative to the number of parameters in order to increase efficiency of the parameter estimates. This is the well-known Generalized Method of Moments (GMM) that has been applied to a wide variety of models in economics. GMM does not necessarily lead to efficient estimates, since the number of elements in the weighting matrix increases too. Asymptotically, however, efficiency increases for the parameters in a GMM estimation (Hansen, 1982). Indeed, Andersen and Sørensen (1994) empirically show that for small sample sizes there is a clear trade-off between the efficiency gains arising from more moments and the deterioration of precision in the weighting matrix.

In terms of efficiency, Jacquier, Polson and Rossi (1994) find that the GMM estimators of Melino and Turnbull (1990) perform worse than their simulation estimators and, in some cases, even to QML. The problem with the GMM estimator for the SV model is that it is sensitive to the number and choice of sample moments, and that some moments depend on the distributional assumptions regarding ϵ_t . Optimal choices of moments are discussed in Gallant, Hsieh and Tauchen (1994), which require simulation estimators of the likelihood and score of an auxiliary model. The analysis of Andersen and Sørensen (1994) tries to alleviate the problem of finding estimates for the elements in the weighting matrix. But, as they show, the improvements are minor, and will not make GMM a serious competitor to simulation estimators.

The Efficient Method of Moments (EMM) of Gallant and Tauchen (1994) tries to fix the efficiency problems that GMM has. This method is related to the indirect inference method of Gourieroux, Monfort, and Renault (1993). An auxiliary model is estimated and the scores of this model with respect to the parameters in the original model determine which moments should be used in a MM analysis. Full efficiency can be achieved if the score of the auxiliary model nests the conditional distribution of the data. Gallant, Hsieh and Tauchen (1994) apply this method to SV models with a nonparametric function as the score generator. Their findings suggest that a nonparametric method would probably do a better job at describing conditional volatility in financial time series than simple lognormal SV models do.

A major disadvantage of estimation methods based on MM in the SV case is that no estimates of the latent volatility processes are obtained. In many financial applications, like option pricing, the interest focuses not so much on the hyperparameters, but on the estimates of the volatility process. However, we can still use MM to obtain suitable starting values for the hyperparameters that can serve as input for simulation-based estimators.

3.1.2 Simulation Estimators

Finding a direct solution to the integral problem posed in the introduction of this section is the aim of simulation algorithms. Several methods have been developed in the last few years; most of them made possible through increasing computer power. Much debate is going on which method to use in even the simplest SV models. The arguments are mainly concerned with computationally efficient ways of approximating the integral (3.3) and finding the hyperparameters. In this field of research the article by Jacquier, Polson and Rossi (1994) (JPR henceforth) describes the state of the art. This is mainly due to the development of a Bayesian Monte Carlo Markov Chain algorithm to estimate the latent

volatilities together with the hyperparameters. JPR also make a thorough comparison of different estimation techniques for SV models (QML, GMM, simulation estimators).

In the empirical analyses later in this chapter we will also compare a number of estimation techniques for SV models. In particular, our focus will be on the simulation estimators that were analyzed in Kim and Shephard (1994). We will present a model that is more general than the model by Kim and Shephard (1994). The simulation estimators in that article rely heavily on a powerful statistical technique called the Gibbs sampler. Basically, the Gibbs sampler is an iterative technique to sample from a joint probability density function. A so-called Gibbs sequence, that is made out of conditional densities, is set up. More precisely, let the triplet (X, Y, Z) be correlated random variables having a joint density function $f(x, y, z)$. Suppose that we want a sample $\{x_t, y_t, z_t\}_{t=1}^N$ from this density. Furthermore, suppose that it is difficult to sample from $f(x, y, z)$ directly, but that the conditional distributions $f(x|y, z)$, $f(y|x, z)$ and $f(z|y, x)$ are in such a form that sampling from them is easier. Given initial values for Y and Z (given by $y^{(0)}$ and $z^{(0)}$) we can set up the following chain

- (i) Set $i = 1$,
- (ii) Draw $x^{(i)}$ from $f(x|y^{(i-1)}, z^{(i-1)})$,
- (iii) Draw $y^{(i)}$ from $f(y|x^{(i)}, z^{(i-1)})$,
- (iv) Draw $z^{(i)}$ from $f(z|y^{(i)}, x^{(i)})$,
- (v) Set $i = i + 1$ and go to (ii).

After a sufficient number of iterations a draw from the joint distribution $f(x, y, z)$ is obtained.² The number of iterations depends on the correlation between successive draws from the chain. Kim and Shephard (1994) show that for SV models this number is relatively low. Repeating the iteration scheme N times gives a proper independent sample from $f(x, y, z)$. Note that the joint distribution is not used in the algorithm. If we know the conditional distributions a sample from $f(x, y, z)$ can be obtained, without even knowing the exact form of $f(x, y, z)$. The order in which the conditional distributions are put in the Gibbs chain does not matter. Especially in cases where samples from high dimensional probability functions are needed, the Gibbs sampler is extremely useful.

²Chan (1993) gives an alternative updating scheme in which the conditional distribution to sample from is chosen randomly.

The Gibbs algorithm is a Markov chain model in which the transition probabilities are given by the conditional densities. Let $\Xi^{(i)} = (x^{(i)}, y^{(i)}, z^{(i)})$ be a realization of the Markov chain. The transition probability is given by

$$p(\Xi^{(i)}, \Xi^{(i+1)}) = f(x^{(i+1)}|y^{(i)}, z^{(i)}) f(y^{(i+1)}|x^{(i+1)}, z^{(i)}) f(z^{(i+1)}|x^{(i+1)}, y^{(i+1)})$$

The Gibbs sampler provides a way in which the distribution of the chain converges to an invariant distribution f . Chan (1993) and Tierney (1991) show that under some regularity conditions the chain converges in distribution to a unique invariant distribution for every initial distribution of the chain. This means that the chain is ergodic: the covariance between the triplets $(X^{(i)}, Y^{(i)}, Z^{(i)})$ and $(X^{(i+n)}, Y^{(i+n)}, Z^{(i+n)})$ goes to zero when $n \rightarrow \infty$. Given the Markov properties the Gibbs sampler belongs to the class of Monte Carlo Markov Chain (MCMC) algorithms.

The Gibbs sampler belongs to the family of so-called data augmentation methods. Another member of this family is the well-known EM algorithm by Dempster, Laird and Rubin (1983). The basic idea of data augmentation methods is that certain characteristics of the data can be estimated in a much simpler way when the observed data is augmented with latent data. An example is the case in which the data is irregularly spaced. The missing observations are modelled as latent variables. In many economic problems, interest lies in the estimation of the latent process itself. Many models in economics are specified in state-space form, *i.e.* the latent process is an integral part of the model (see Harvey, 1989). Hyperparameters can be estimated iteratively by computing the expectation over the augmented data and maximizing the resulting function with respect to the hyperparameters. The EM algorithm provides the maximum of the objective function, whereas the Gibbs sampler and other data augmentation methods give a complete probability density function of the relevant random variables or statistics. See Tanner (1991) for an extensive comparison of the different methods.

In the application to exchange rates in the next section we deal with the joint estimation of the hyperparameters and the latent variables of a general state-space with a MCMC algorithm. The simulation method makes use of the linearization of the state-space, and approximates the resulting log chi-squared distribution by a mixture of normals. This algorithm makes full use of the Gibbs sampler. Although MCMC algorithms were originally designed to be used in Bayesian analysis, we discuss a maximum likelihood estimation technique that also uses MCMC. This technique is called Simulated Expectation Maximization (SIEM) (Kim and Shephard, 1994).

JPR combine a Gibbs sampler with a general Metropolis-Hastings (MH) algorithm to obtain the marginal posterior densities of the parameters in (3.1) and (3.2), and also the

exact posterior distribution of the variance series given the observed data $\ln s_t^2$. In fact the Gibbs sampler is a special case of the MH algorithm. The MH algorithm has a more general transition probability function than the Gibbs sampler (Geweke, 1994). The basic idea of JPR is to construct a posterior distribution of the two unknown quantities in the SV model: the parameter vector θ and the vector of latent volatilities H_T . The resulting posterior distribution is

$$f(H_T, \theta | S_T) \propto f(S_T | H_T) f(H_T | \theta) f(\theta),$$

where $f(S_T | H_T)$ is the likelihood part and $f(H_T | \theta)$ and $f(\theta)$ form a joint prior over (H_T, θ) . This illustrates the idea of data augmentation, that we referred to above. The vector of parameters is augmented by the latent vector H_T , thus forming a joint vector of quantities to be estimated. We are interested in the marginals $f(H_T | S_T, \theta)$ and $f(\theta | S_T, H_T)$, which can be found by constructing a MCMC chain that cycles between the two densities. The difficulty that now arises concerns the sampling from $f(H_T | S_T, \theta)$ for which JPR set up a MH exercise. Instead of sampling from the multivariate density $f(H_T | S_T, \theta)$ directly, JPR break this density up into T conditional distributions $f(h_t | h_{-t}, S_T, \theta)$ ($t = 1, \dots, T$), where h_{-t} is a $T-1$ vector in which the element $\{h_t\}$ is skipped. For every t a MH chain is set up. JPR show that their Markov chain, called *cyclic dependence Metropolis* chain, converges to an invariant distribution.

Recently, several other simulation methods have been developed to deal with the multiple integration problem in (3.3). These methods also heavily rely on simulation techniques. Danielsson (1994) and Danielsson and Richard (1993) develop an importance sampling technique to estimate the value of the integral. This method is still very computer intensive and the resulting algorithms cannot easily be generalized to more general SV models. Furthermore, extensions to multivariate models seem to be much more involved than is the case for algorithms based on MCMC.

3.2 Mixture Model

In this section we will discuss an alternative way – different from JPR – to estimate the lognormal SV model. Kim and Shephard (1994) (KS hereafter) show that the MCMC in the JPR model typically takes more time to converge than their method. We will take the KS analysis as a starting point for investigating the simple SV model through simulated maximum likelihood methods. Let S_t be a bilateral exchange rate, and define $s_t = \Delta \ln S_t$. Assuming that the change in the logarithm of the exchange rate is unpredictable, the

standard stochastic volatility model can be written as³

$$s_t = \exp(h_t/2)\epsilon_t, \quad (3.5)$$

$$h_t = \beta + \phi(h_{t-1} - \beta) + \eta_t, \quad (3.6)$$

where $\exp(h_t)$ is the variance of s_t at time t , and where the innovations ϵ_t and η_t have mean zero, with variances equal to one and σ_η^2 respectively. The usual assumption is that ϵ_t and η_t are normally distributed. The exchange rate obtains its fat tailed distribution by the mixing of ϵ_t and $\exp(h_t/2)$ as was seen in Chapter 2. The estimation has been the main obstacle for application of this type of model. From a computational viewpoint specification (3.5) with normality for ϵ_t is inconvenient, since the likelihood function can only be written in integral form (3.3).

As was noted before, the most straightforward way to estimate the SV model is QML, which starts by transforming the measurement equation (3.5). Let $y_t = \ln s_t^2$, then (3.5) can be written in the linear form

$$y_t = h_t + \xi_t, \quad (3.7)$$

where $\xi_t = \ln \epsilon_t^2$. QML estimates of the parameters can be obtained by standard numerical optimization techniques, since the full likelihood becomes Gaussian, if we assume that the distribution of ξ_t is normal. As we have seen in Section 3.1.1 the QML estimator is not efficient, since the transformed error term ξ_t will be extremely skewed to the left, if the underlying ϵ_t is normal. Not only can this result in inefficient parameter estimates, but the standard Kalman smoother might also produce poor estimates for the state variable h_t conditional on the parameters of the process.

Shephard (1994) and KS retain the convenient linear form of the state-space form, and approximate the logarithmic chi-squared distribution of ξ_t by a prespecified mixture of seven normal distributions. Shephard (1994) describes efficient algorithms for statistical inference in this class of what he calls partial non-Gaussian state-space models, and provides examples of how to take advantage of the linear and Gaussian parts. The mixture is specified as

$$\begin{aligned} \xi_t &= \xi(z_t), \quad z_t = 1, \dots, M, \\ \xi(i) &\sim N(\mu_i, \omega_i^2) \quad \text{with probability } p_i, \end{aligned}$$

where M is the number of mixtures. Using the mixture model the value of the likelihood can be estimated using the multimove Gibbs sampler methodology from Section 3.1.2. Conditional on a time series of indicators z_t ($t = 1, \dots, T$), the standard Kalman recursions

³The unpredictability of exchange rates (See *e.g.* De Vries, 1992, and the references therein) is under scrutiny again. See Canova (1993) and Mark (1995).

can be used to simulate the states h_t , because the measurement error has a normal distribution now. We collect all the mixture indicators z_t in a vector Z_T . Given new states we can sample new mixture indicators again. Collect the y_t ($t = 1, \dots, T$) in the vector Y_T . The Markov chain consists of the following steps:

1. Provide some initial series for the mixture indicators: $Z_T^{(0)}$,
2. Set $i = 1$,
3. Draw the state vector $H_T^{(i)} = (h_1, \dots, h_T)$ from the density $f(H_T | Y_T, Z_T^{(i-1)}; \theta)$,
4. Draw new mixture indicators $Z_T^{(i)}$ from the density $\Pr[Z_T | Y_T, H_T^{(i)}; \theta]$,
5. Set $i := i + 1$ and go to 3.

The vector θ contains the parameters β , ϕ and σ_η^2 . The chain provides us with a sample $(H_T^{(i)}, Z_T^{(i)})$ from the joint distribution $f(H_T, Z_T | Y_T; \theta)$. Given this sample we can approximate any statistic that is a function of H_T and/or Z_T .

The drawing of new states in step 3 of the chain proceeds in two steps. Note that the simulated states are conditional on all the data Y_T , meaning that we need to simulate *smoothed* states. First a normal Kalman filter is run, in order to get predicted states. For ease of exposition move the parameter β from the state transition equation to the measurement equation. This leads to the equivalent state-space form

$$\begin{aligned} y_t &= h_t + \beta + \xi_t, \\ h_t &= \phi h_{t-1} + \eta_t. \end{aligned}$$

The filtered states \hat{h}_t and its corresponding conditional variance \hat{p}_t^2 are given by

$$\begin{aligned} \hat{h}_t &= \phi \hat{h}_{t-1} + \frac{\hat{p}_t^2}{\omega^2(z_t)} [y_t - \phi \hat{h}_{t-1} - \beta - \mu(z_t)], \\ \hat{p}_t^2 &= \frac{\omega^2(z_t)(\phi^2 \hat{p}_{t-1}^2 + \sigma_\eta^2)}{\phi^2 \hat{p}_{t-1}^2 + \sigma_\eta^2 + \omega^2(z_t)}, \end{aligned}$$

where we have suppressed the iteration number i . We apply initial values $\hat{h}_0 = 0$ and $\hat{p}_0^2 = \sigma_\eta^2 / (1 - \phi^2)$.

In the second part states are drawn from the results of the Kalman smoothing recursions. Going backwards from period $t = T$ to period $t = 1$, the simulation smoother draws states (h_t^s) from the normal distribution

$$h_t^s \sim N \left[\hat{h}_t + \frac{\phi \hat{p}_t^2}{\phi^2 \hat{p}_t^2 + \sigma_\eta^2} (h_{t+1}^s - \phi \hat{h}_t), \frac{\sigma_\eta^2}{\phi^2 \hat{p}_t^2 + \sigma_\eta^2} \right].$$

The final state h_T^s is drawn from $N(\hat{h}_T, \hat{p}_T^2)$. The mean and variance of the normal distribution are given by the fixed-interval smoother (Harvey, 1989).

New mixture indicators (step 4 of the chain) given a state vector H_T can be found as follows. The distribution $\Pr[Z_T|Y_T, H_T^{(i)}; \theta]$ is given by the density

$$\Pr[Z_T|Y_T, H_T^{(i)}; \theta] = \prod_{t=1}^T f(y_t|z_t, h_t^s) \Pr(z_t).$$

Note that each density in the product is multinomial with M posterior probabilities

$$\Pr[z_t|y_t, h_t^s] \propto \frac{\Pr(z_t)}{\omega(z_t)} \exp \left(-\frac{1}{2} \left[\frac{y_t - h_t^s - \beta - \mu(z_t)}{\omega(z_t)} \right]^2 \right) \quad z_t = 1, \dots, M. \quad (3.8)$$

Since this is a discrete probability distribution the constant of proportionality can be found very easily by summing the right-hand side of (3.8) for every $z_t = 1, \dots, M$. The simulation is very simple since for every t only one uniformly distributed random variable is needed. Comparison of this random variable with the cumulative distribution implied by (3.8) gives the new mixture indicator. Thus, the sampling of mixture indicators z_t is nothing more than a posterior odds calculation using the posteriors given in (3.8).

KS show that the number of rounds needed in the Gibbs sampler to obtain good estimates of the states is very low, i.e. the Gibbs sampler converges fast to the invariant distribution of the underlying Markov chain. This is due to the low correlation between the draws made in the chain. Thus, a relatively small "burn-in" period is needed. As KS show, this contrasts markedly with the Metropolis-Hastings algorithm of JPR, which takes much longer to converge.

KS use a variant of the EM algorithm to maximize the likelihood function. Originally developed by Dempster, Laird and Rubin (1977), the EM algorithm finds the mode of the posterior distribution of the parameters. Every iteration in the algorithm consists of two steps: an expectations step (E-step) and a maximization step (M-step). In the E-step the latent variables are integrated out from the likelihood. The resulting quantity is then maximized with respect to the parameters. A very convenient property of the EM algorithm is that consecutive parameter iterates increase the likelihood with probability one.

The EM algorithm in our case is motivated by decomposing the likelihood as in

$$\ln f(Y_T; \theta) = \ln f(Y_T|H_T, Z_T; \theta) + \ln f(H_T, Z_T) - \ln f(H_T, Z_T|Y_T; \theta).$$

From Dempster, Laird and Rubin (1977) and Tanner (1991) we know that the maximization

of the likelihood is given by

$$\theta^* = \arg \max_{\theta} \frac{1}{N} \sum_{i=1}^N \ln f(Y_T | H_T^{(i)}, Z_T^{(i)}; \theta), \quad (3.9)$$

where the superscript denotes the i th drawing from the Gibbs sampler. Performing the maximization in (3.9), the new parameters ϕ , σ_{η}^2 and β can be found as

$$\begin{aligned} \hat{\phi}^* &= \frac{\sum_{j=1}^N \sum_{t=2}^T h_t^{(j)} h_{t-1}^{(j)}}{\sum_{j=1}^N \sum_{t=2}^T (h_{t-1}^{(j)})^2}, \\ \hat{\sigma}_{\eta}^{2,*} &= \frac{1}{(T-1)N} \sum_{j=1}^N \sum_{t=2}^T (h_t^{(j)} - \hat{\phi} h_{t-1}^{(j)})^2, \\ \beta^* &= \frac{1}{NT} \sum_{j=1}^N \sum_{t=1}^T (y_t - h_t^{(j)} - \mu(z_t^{(j)})). \end{aligned}$$

Note that only the new estimate for β depends directly on the mixture distributions. The new estimates are used again in a next round of the Gibbs sampler to find new simulated states and mixture indicators until convergence of the objective function in (3.9) is obtained.

To find a value for the likelihood function it is useful to employ the following decomposition

$$\ln L(\theta | Y_T) = \ln \Pr(Z_T) + \ln f(Y_T | Z_T; \theta) - \ln \Pr(Z_T | Y_T; \theta)$$

Only the last term on the right hand requires simulation from the multimove Gibbs sampler, since the first term is directly given by the mixture indicators and the second term can be found by the Kalman filter. As KS show we can use draws from the distribution $\Pr[Z_T | Y_T, H_T^{(i)}; \theta]$ to estimate the last term.

The main feature of the data are the so-called inliers, *i.e.* observations with almost zero change for which the log transformation generates large negative outliers. Even as few as two normals can model the empirically observed (and also expected) negative skewness of ξ_t . A mixture with the first normal centered around zero, covering the bulk of the data, and the other centered around a large negative mean to accommodate the outliers, seems a good candidate for the distribution of ξ_t . Its main drawback is that such a distribution is likely to be bimodal. After some experimentation we therefore settled for a mixture of three normals with weights $p = (0.70, 0.25, 0.05)$. The middle element in the mixture provides a smooth blending to a unimodal density. We chose to fix the mixture weights at predetermined constants, because it turned out, through experimentation, that a relatively low number of degrees of freedom is sufficient for the mixture of normals to approach the log chi-squared distribution.

TABLE 3.1: Mixture Parameters

| i | weight | μ_i | σ_i |
|---|--------|---------|------------|
| 1 | 0.70 | -0.2172 | 1.1052 |
| 2 | 0.25 | -3.0461 | 1.5705 |
| 3 | 0.05 | -6.4818 | 3.0002 |

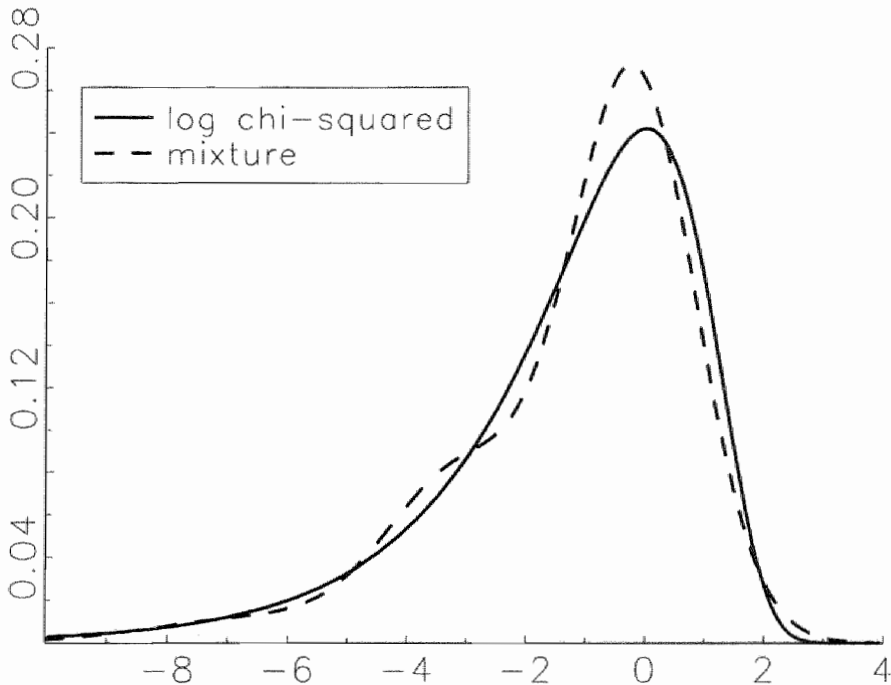


FIGURE 3.3: Densities of Measurement Error

A mixture of three normals is not a serious limitation in case the error distribution is log chi-squared. We performed a small simulation study to estimate the optimal parameters for data generated by a log chi-squared. The resulting mixture distribution parameters are given in Table 3.1. Figure 3.3 plots the mixture distribution together with the actual log chi-squared distribution. The main difference is the slightly higher mode of the mixture distribution. It looks like a three element mixture strikes a balance between flexibility and the number of parameters.

We apply the "Simulated EM" algorithm to a number of exchange rates in the next section. The number of mixtures is set to three ($M=3$). In the initial stages of the EM algorithm we used 640 replications from the Gibbs sampler, of which every fifth was used in the maximization (3.9) to diminish possible serial correlation in the draws. After 50 iterations of EM, we increased the number of replications to 5,120, again using only every fifth replication. Evidence in Kim and Shephard (1994) for SV models with parameters that correspond with the parameters typically found for financial time series, suggests that a relatively low number of replications is needed for the Gibbs sampler to converge. In fact, a number as low as 10 seems to be enough. The simulated EM with a mixture of normals that replicates the log chi-square distribution of the measurement errors is called "SIEM1" to distinguish it from an extension that we will discuss later in this section.

Although the transformation from ξ_t back to ϵ_t is well defined, it is very sensitive to exact specification of especially the right tail of the density of ξ_t . For example, under the QML assumption $\xi_t \sim N(-1.27, \pi^2/2)$, the implied kurtosis of ϵ_t is equal to $\exp(\pi^2/2) = 139$ instead of 3. For the mixture in Table 3.1 the implied kurtosis is 15. Increasing the number of elements in the mixture does not seem to make much difference. The seven element mixture reported in KS also implies a kurtosis of 14 for the normal random variable ϵ_t , although their mixture excellently describes the first four moments of ξ_t . In general the tails of ϵ_t become fatter the larger the variance of ξ_t , *ceteris paribus*, but given the sensitivity it is hard to draw firm conclusions.

The second simulated EM algorithm that we study in this chapter considers the mean and variance parameters in the normal mixture as parameters. We will refer to this method by the abbreviation "SIEM2". Since the unconditional mean of the measurement error ξ_t is a free parameter in the mixture distribution, the constant β becomes unidentified. We need to maximize the EM objective function (3.9) over the parameter vector $\theta = (\phi, \sigma_\eta^2, \mu_i, \omega_i^2, i = 1, 2, 3)'$. Since ϕ and σ_η^2 are not directly affected by the measurement error distributions the analytical forms of the optimal solutions are not affected either. The optimal estimates for μ_i and ω_i^2 are found to be

$$\begin{aligned}\hat{\mu}_i &= \frac{\sum_{j=1}^N \sum_{t \in I_i^{(j)}} (y_t - h_t^{(j)})}{\sum_{j=1}^N T_i^{(j)}}, \\ \hat{\omega}_i^2 &= \frac{\sum_{j=1}^N \sum_{t \in I_i^{(j)}} (y_t - h_t^{(j)} - \hat{\mu}_i)^2}{\sum_{j=1}^N T_i^{(j)}},\end{aligned}$$

where $I_i^{(j)} = \{t : z_t^{(j)} = i\}$. The algorithm proceeds with a Gibbs sampler round with the new parameters until convergence is reached.

The final estimation method we consider is a Bayesian Gibbs sampling algorithm, in which

we cycle through simulating the states, the mixture indicators, and the parameter vector θ . A full description of the algorithm is available in Appendix 3.A. In the Bayesian procedure we assume very diffuse proper conditionally conjugate priors. The prior means of μ_i and ω_i are set at the values of the approximating mixture in Table 3.1. The prior mean of ϕ is equal to one. The prior variances of the μ_i 's and ϕ are equal to 1000, while the prior degrees of freedom in the inverted gamma priors for ω_i is equal to 5. We need proper priors in order to ensure a proper posterior. With an improper prior on ω_i the posterior does not exist, and it is this pathology that we must avoid. Simulation from the conditional density of θ given the states and the data works well, since the model is linear. The results in the empirical applications are based on a single run of the Gibbs sampler with 50,000 iterations, of which every tenth iteration is used in the empirical analyses. We performed an extra number of 200 iterations to start the Gibbs sampler, which were not used in the analysis. The number of 5,000 iterations that were used in the empirical analysis was high enough to provide posterior distributions that do not change significantly if the results from an extra number of iterations is added.

3.3 Parameter Estimates

The data consist of weekly observations of the bilateral exchange rates among the major currencies (U.S. dollar, U.K. pound sterling, Japanese yen, and German mark). The sample period is January 3, 1973 until February 9, 1994 (1102 observations). The original data are denominated in pound sterling; cross rates have been constructed. All data are taken from DATASTREAM and are sampled on Wednesdays. If Wednesday is a holiday Thursday is taken for all series. The data are transformed to $y_t = \ln[(\Delta \ln S_t - \bar{s})^2]$, where \bar{s} is the sample mean of $\Delta \ln S_t$. The subtraction avoids the problems that might occur due to zero currency returns. We chose to work with weekly data – instead of data that is sampled at a higher frequency – to circumvent the effects of possible seasonality in high-frequency data.

Table 3.2 provides summary statistics. The main features of the transformed data are the negative skewness, and the persistent autocorrelations. Negative skewness is implied by the log transformation, and is related to the "inlier" problem. The variance of the transformed data is much higher than that of a log-chisquared (which is equal to $\pi^2/2$), so that either the volatility series h_t is itself highly volatile or normality of the measurement error in (3.5) is violated.

We include all six possible bilateral rates, because in this way we can model all variances

TABLE 3.2: Summary statistics of $y_t = \ln[(\Delta \ln S_t - \bar{s})^2]$

| | JP/US | GE/US | UK/US | GE/JP | UK/JP | GE/UK |
|------------------------|--------|--------|--------|--------|--------|--------|
| mean | -1.203 | -0.770 | -1.003 | -0.926 | -0.979 | -1.494 |
| variance | 6.555 | 5.954 | 6.383 | 5.318 | 6.461 | 5.817 |
| skewness | -0.971 | -1.202 | -0.968 | -1.257 | -1.278 | -0.904 |
| kurtosis | 1.225 | 1.022 | 0.895 | 2.850 | 2.731 | 1.181 |
| minimum | -13.03 | -13.15 | -11.80 | -15.04 | -14.93 | -11.83 |
| maximum | 4.871 | 4.129 | 4.310 | 3.613 | 4.525 | 3.853 |
| normality | 246.7* | 452.7* | 208.7* | 662.6* | 641.8* | 214.0* |
| LB(10) | 219.2* | 65.65* | 277.7* | 34.33* | 34.90* | 67.95* |
| LB(20) | 383.6* | 127.1* | 413.9* | 54.87* | 69.24* | 94.10* |
| Autocorrelations y_t | | | | | | |
| 1 | 0.184 | 0.080 | 0.172 | 0.078 | 0.074 | 0.132 |
| 2 | 0.142 | 0.098 | 0.170 | 0.071 | 0.081 | 0.096 |
| 3 | 0.163 | 0.115 | 0.152 | 0.105 | 0.064 | 0.090 |
| 4 | 0.167 | 0.055 | 0.202 | 0.047 | 0.077 | 0.114 |
| 5 | 0.127 | 0.096 | 0.123 | 0.063 | 0.016 | 0.035 |
| 10 | 0.104 | 0.027 | 0.129 | -0.012 | 0.019 | 0.053 |
| 20 | 0.087 | 0.062 | 0.073 | 0.097 | 0.040 | 0.002 |

Notes: The kurtosis is measured as excess kurtosis. *Normality* is the Jarque-Bera test for normality. *LB(m)* is the Ljung-Box test taking m autocorrelations. An asterisk '*' denotes significance at the 1% level.

and all possible covariances. For example, the covariance between the yen/pound and mark/pound can be written as the identity

$$\text{Cov}(\Delta \ln S^{JP/UK}, \Delta \ln S^{GE/UK}) =$$

$$\frac{1}{2} [\text{Var}(\Delta \ln S^{JP/UK}) + \text{Var}(\Delta \ln S^{GE/UK}) - \text{Var}(\Delta \ln S^{JP/GE})],$$

which shows that the covariance between two pound denominated exchange rates can be obtained through the variance of the cross rate.

The stochastic volatility model with an AR(1) specification for the volatility will be estimated by simulation methods. We compare three different estimators. The first is the simulated EM algorithm of KS, which converges to the maximum likelihood estimator. With this estimator the parameters of the mixture are fully specified *a priori* to mimic the log-chisquared distribution (SIEM1). An alternative is the simulated EM method with

free parameters in the three elements of the mixture (SIEM2). The SIEM2 model allows flexibility in the error distribution, and can accommodate fat-tailed distributions of the underlying exchange rate innovation ϵ_t . We monitor convergence of the methods by graphical inspection of the EM iterates.

The relatively straightforward way to compute the likelihood value in the SIEM algorithms allows us to get standard errors for the parameters. Since both SIEM1 and QML are nested within the general SIEM2, these two specifications can be tested by a likelihood ratio test. Note that QML is nested within SIEM2 by equalizing the means and variances in each conditional normal distribution of the mixture. We can numerically differentiate the likelihood under the null hypothesis and consequently find both the gradient and Hessian. Furthermore LM tests can be performed.

Tables 3.3 and 3.4 summarize the parameter estimation results of the four estimators/models: QML, simulated EM with fixed mixture (SIEM1), simulated EM with flexible mixture (SIEM2), and the Bayesian Gibbs sampler (BAYES).

In general the parameter estimates are similar, both across currencies as well as estimators. This is as expected due to the large sample of more than twenty years of weekly data and the tight parameterization. The high value of ϕ implies persistent logarithmic volatility series. However, when the measurement error is kept fixed (QML and SIEM1) the persistence is much lower for some series, notably the pound/yen and pound/mark series. This might be due to outliers, which are harder to accommodate by QML and the fixed mixture. Another explanation is the importance of the identifying restriction $\omega_i^2 \equiv \omega^2 = \pi^2/2$ ($i = 1, 2, 3$) in QML estimation, as we will discuss later in this section. In general the persistence of the volatility rises when the measurement error distribution is flexible. Also, when ϕ is small, the estimate of the volatility innovation variance σ_η^2 increases. The estimates of ϕ are never significantly different from unity using Dickey-Fuller critical values. The Bayesian posterior means of ϕ are somewhat lower than the ML estimates in SIEM2. This is reflected in the marginal posterior densities of ϕ in Figure 3.4, which are all skewed to the left. One way to increase the posterior mean is by adopting a different prior, for instance one that is proportional to $(1 - \phi^2)^{-1/2}$. Such a prior would implicitly arise, if it is assumed that the initial condition in the state vector has variance proportional to $(1 - \phi^2)^{-1}$ (see Schotman, 1994 and KS).

Since the SIEM2 model nests both the QML and the SIEM1 model, likelihood ratio tests can be computed. As reported in Table 3.5 both the normal (QML) and the log-chisquared distributions can be firmly rejected in every case against the free six parameter mixture. The QML and SIEM1 estimators restrict the second and higher moments, while the first

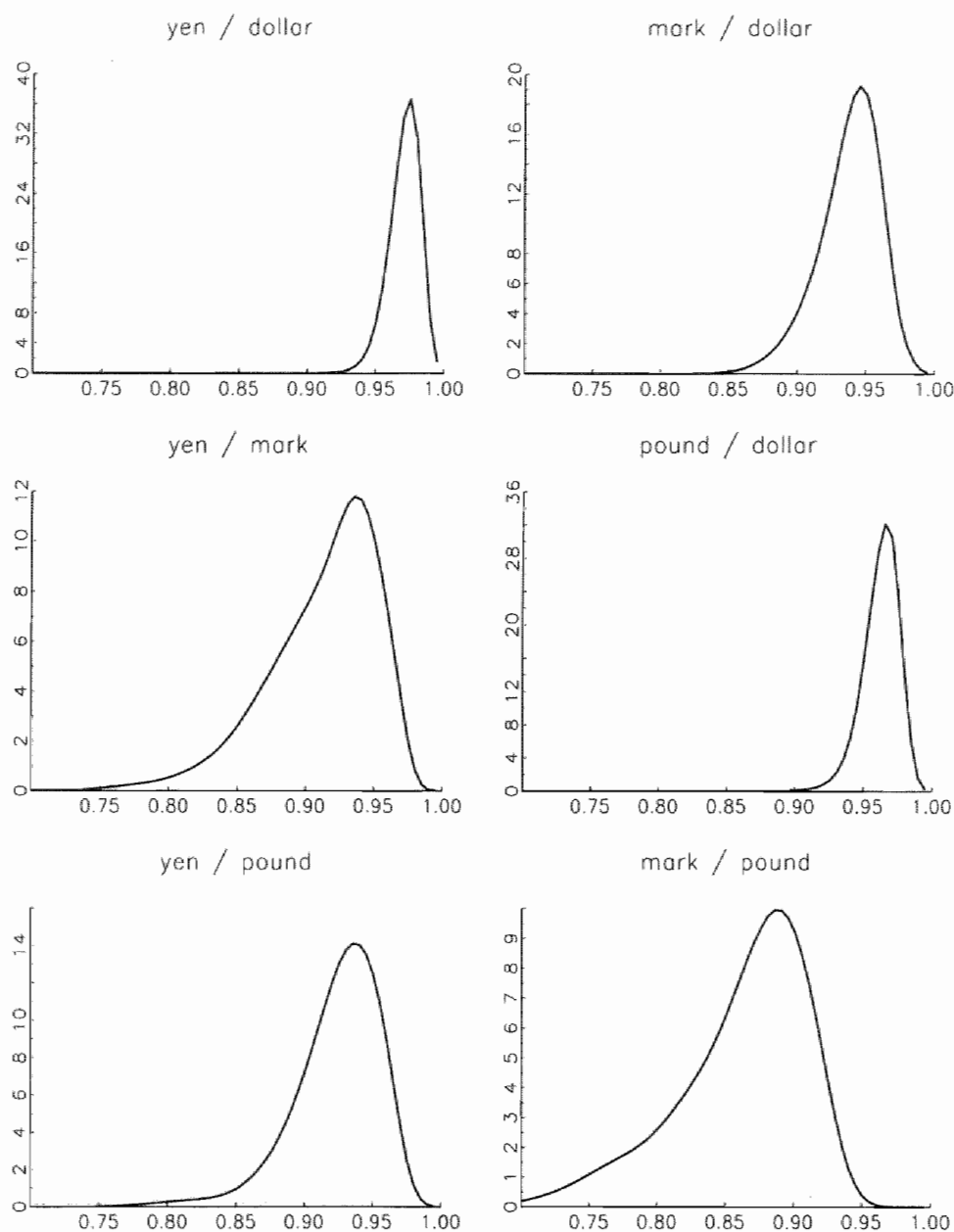
FIGURE 3.4: Posterior Densities of ϕ

TABLE 3.3: Parameter Estimates of Stochastic Volatility Model

| | $y_t = h_t + \xi_t,$ | | | $\xi_t \sim \sum_i p_i N(\mu_i, \omega_i^2)$ | | |
|---|--|-------------------|-------------------|--|-------------------|-------------------|
| | $h_t = \beta(1 - \phi) + \phi h_{t-1} + \eta_t,$ | | | $\eta_t \sim N(0, \sigma_\eta^2)$ | | |
| | JP/US | GE/US | UK/US | GE/JP | UK/JP | GE/UK |
| Quasi Maximum Likelihood (QML) | | | | | | |
| ϕ | 0.976 (0.015) | 0.967 (0.027) | 0.960 (0.015) | 0.985 (0.009) | 0.653 (0.391) | 0.884 (0.061) |
| β | 0.050 (0.211) | 0.482 (0.199) | 0.238 (0.258) | 0.369 (0.190) | 0.292 (0.101) | -0.226 (0.130) |
| σ_η | 0.225 (0.081) | 0.198 (0.097) | 0.311 (0.055) | 0.092 (0.027) | 0.842 (0.612) | 0.419 (0.142) |
| Simulated EM, fixed mixture (SIEM1) | | | | | | |
| ϕ | 0.878 (0.019) | 0.928 (0.018) | 0.952 (0.010) | 0.921 (0.021) | 0.584 (0.048) | 0.768 (0.034) |
| β | 0.150 (0.138) | 0.591 (0.123) | 0.395 (0.200) | 0.362 (0.111) | 0.349 (0.068) | -0.194 (0.090) |
| σ_η | 0.558 (0.040) | 0.293 (0.036) | 0.330 (0.029) | 0.285 (0.039) | 0.822 (0.055) | 0.650 (0.050) |
| Simulated EM, flexible mixture (SIEM2) | | | | | | |
| ϕ | 0.979 (0.008) | 0.975 (0.010) | 0.967 (0.009) | 0.954 (0.016) | 0.957 (0.016) | 0.930 (0.020) |
| σ_η | 0.190 (0.029) | 0.148 (0.027) | 0.251 (0.028) | 0.200 (0.035) | 0.190 (0.040) | 0.285 (0.042) |
| μ_1 | -0.042 (0.266) | 0.410 (0.176) | 0.205 (0.227) | 0.102 (0.137) | 0.133 (0.140) | -0.353 (0.129) |
| μ_2 | -3.122 (0.282) | -2.724 (0.199) | -2.190 (0.230) | -2.547 (0.156) | -2.736 (0.174) | -3.441 (0.144) |
| μ_3 | -6.651 (0.429) | -6.877 (0.416) | -6.561 (0.282) | -6.819 (0.344) | -7.355 (0.447) | -6.048 (0.244) |
| ω_1^2 | 1.755 (0.097) | 1.281 (0.075) | 1.417 (0.081) | 1.277 (0.080) | 1.775 (0.105) | 1.512 (0.090) |
| ω_2^2 | 2.578 (0.230) | 2.759 (0.234) | 1.537 (0.133) | 2.040 (0.182) | 3.439 (0.297) | 1.436 (0.144) |
| ω_3^2 | 6.134 (1.214) | 7.981 (1.520) | 2.005 (0.377) | 5.865 (0.865) | 10.085 (1.840) | 2.385 (0.403) |

Notes: Numbers in parenthesis are robust standard errors. Parameters in fixed mixture are as given in Table 3.1. The mixture weights were set to (0.70, 0.25, 0.05). In the QML model ξ_t has mean -1.27 and variance $\pi^2/2$. In the SIEM2 model β is restricted to zero for identification.

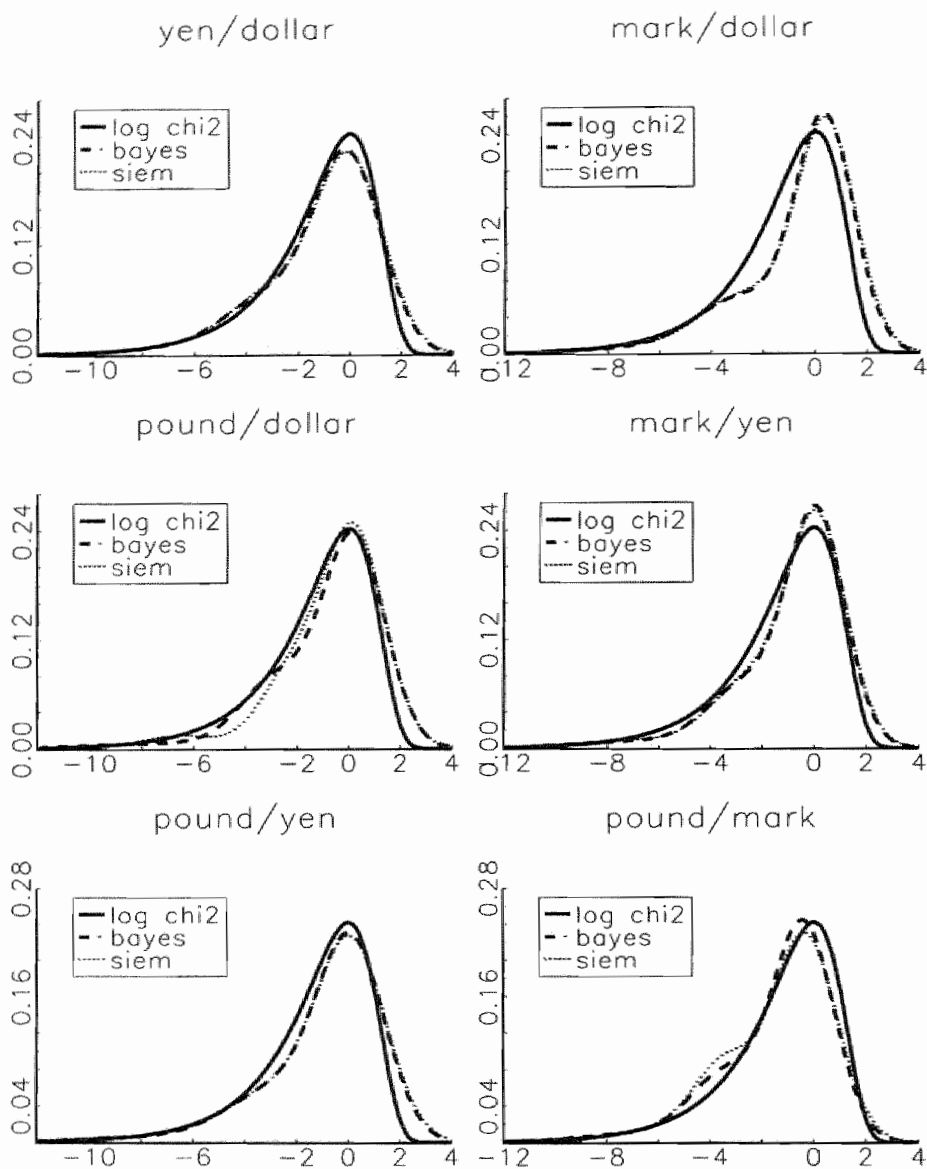


FIGURE 3.5: Implied Measurement Error Densities

TABLE 3.4: Bayesian Posterior Moments

| | | | | | | |
|---------------|--------------------------------|------------------|--|------------------|------------------|------------------|
| | $y_t = h_t + \xi_t,$ | | $\xi_t \sim \sum_i p_i N(\mu_i, \omega_i^2)$ | | | |
| | $h_t = \phi h_{t-1} + \eta_t,$ | | $\eta_t \sim N(0, \sigma_\eta^2)$ | | | |
| | JP/US | GE/US | UK/US | GE/JP | UK/JP | GE/UK |
| ϕ | 0.971 (0.013) | 0.938 (0.022) | 0.963 (0.014) | 0.912 (0.041) | 0.924 (0.033) | 0.862 (0.050) |
| σ_η | 0.274 (0.046) | 0.267 (0.047) | 0.285 (0.044) | 0.300 (0.082) | 0.270 (0.059) | 0.454 (0.101) |
| μ_1 | -0.091 | 0.359 | 0.188 | 0.092 | 0.113 | -0.390 |
| μ_2 | -3.137 | -2.778 | -2.800 | -2.587 | -2.796 | -3.415 |
| μ_3 | -6.389 | -5.906 | -6.727 | -6.574 | -6.807 | -6.883 |
| ω_1 | 1.315 | 1.113 | 1.208 | 1.108 | 1.305 | 1.174 |
| ω_2 | 1.691 | 1.754 | 1.366 | 1.474 | 1.847 | 1.374 |
| ω_3 | 2.706 | 2.979 | 3.182 | 2.534 | 3.220 | 2.146 |

Notes: Posterior means and standard deviations (in parentheses) of the parameters are based on a run of 5000 simulations from the Gibbs sampler. The weights of the mixture are (0.70, 0.25, 0.05).

TABLE 3.5: Likelihood Ratio Tests

| | | | | | | |
|-------------------------|-------|-------|-------|-------|-------|-------|
| | JP/US | GE/US | UK/US | GE/JP | UK/JP | GE/UK |
| QML (normal) | 153.0 | 263.6 | 134.4 | 227.0 | 244.2 | 109.0 |
| SIEM1 ($\ln(\chi^2)$) | 153.4 | 76.0 | 42.4 | 50.4 | 234.4 | 108.2 |

Notes: The likelihood ratio statistic has a $\chi^2(5)$ distribution for both hypotheses. The 5% and 1% critical values are 11.1 and 15.1 respectively.

moment is unrestricted (because of β). Since the tests solely involve the shape of the measurement error distribution, it is of interest to compare the first few moments of the mixture distributions. The formulas for the moments of a mixture of normals are listed in Appendix 3.B. Table 3.6 reports the implied moments of the measurement errors. Overall the implied variance estimates of the flexible mixture are larger than those of the log-chisquared distribution. The high variance of the measurement error ξ_t suggests that the distribution of the ϵ_t in (3.5) has fatter tails than the normal. Gallant, Hsieh and Tauchen (1994) reach a similar conclusion in an application with stock price data. The BAYES moments are computed by averaging over the conditional moments in a run of the Gibbs

TABLE 3.6: Moments of Measurement Error Distribution

| | JP/US | GE/US | UK/US | GE/JP | UK/JP | GE/UK | SIEM1 | QML |
|------------------------|--------|--------|--------|--------|--------|--------|--------|-------|
| Variance | | | | | | | | |
| SIEM2 | 5.524 | 5.778 | 5.042 | 4.830 | 6.276 | 5.289 | 4.843 | 4.935 |
| BAYES | 5.511 | 5.349 | 5.087 | 4.767 | 5.987 | 4.937 | | |
| Skewness | | | | | | | | |
| SIEM2 | -1.094 | -1.500 | -1.082 | -1.482 | -1.512 | -0.681 | -1.478 | 0 |
| BAYES | -1.133 | -1.386 | -1.106 | -1.475 | -1.436 | -1.142 | | |
| Excess Kurtosis | | | | | | | | |
| SIEM2 | 1.847 | 3.244 | 2.129 | 3.454 | 3.776 | 0.842 | 3.514 | 0 |
| BAYES | 2.045 | 2.741 | 1.608 | 3.422 | 3.457 | 1.860 | | |

Notes: The SIEM2 moments are computed using the parameter estimates in Table 3.3. The SIEM1 moments are the same for all exchange rates and were fixed in advance using the parameters in Table 3.1. The BAYES moments are computed as the average of the conditional moments over the iterations of the Gibbs sampler. Skewness and kurtosis are normalized, and the reported kurtosis is in excess of the normal kurtosis.

sampler. In many cases the BAYES moments are close to the maximum likelihood results, although the implied variances are somewhat smaller. There is no discernible pattern of differences for the higher order moments for the six series. Figure 3.5 shows a plot of the estimated measurement densities for ξ_t . The Bayesian densities are obtained as the average of the conditional densities over a run of the Gibbs sampler. Note that the BAYES densities are very close to the SIEM densities.

In Table 3.7 we present some selected moment statistics of the actual data in order to further compare the four models. The main conclusion from the table is that the four estimators/models have widely varying fits for the different moments of the log transformed data. For instance, QML underestimates the variance of $\ln s_t^2$ consistently. Major differences between the estimators show up for the excess kurtosis, where the BAYES estimator implies much less kurtosis than is actually in the data.

The implied autocorrelations are curious. For example, QML and SIEM1 provide the lowest estimates of ϕ for the pound/yen volatility in Table 3.3. But even so they imply the highest first order autocorrelations for the data. The explanation is that the measurement error variance has been fixed for QML and SIEM1, and is much lower than the unrestricted estimate of SIEM2. Since y_t follows an ARMA(1,1) process, parameterized by ϕ , σ_η^2 and ω^2 , restricting $\omega^2 = \pi^2/2$, as in QML and SIEM1, will affect the estimates of the remaining

TABLE 3.7: Data and Implied Model Moments

| | JP/US | GE/US | UK/US | GE/JP | UK/JP | GE/UK |
|-------------------------------------|--------|--------|--------|--------|--------|--------|
| Mean | | | | | | |
| $\ln(s^2)$ | -1.203 | -0.770 | -1.003 | -0.926 | -0.979 | -1.494 |
| QML | -1.220 | -0.788 | -1.032 | -0.901 | -0.978 | -1.496 |
| SIEM1 | -1.088 | -0.647 | -0.843 | -0.876 | -0.889 | -1.432 |
| SIEM2 | -1.142 | -0.738 | -0.732 | -0.906 | -0.959 | -1.410 |
| BAYES | -1.167 | -0.738 | -0.905 | -0.911 | -0.960 | -1.471 |
| Variance | | | | | | |
| $\ln(s^2)$ | 6.555 | 5.954 | 6.383 | 5.318 | 6.461 | 5.817 |
| QML | 6.002 | 5.539 | 6.168 | 5.219 | 6.171 | 5.738 |
| SIEM1 | 6.201 | 5.463 | 6.007 | 5.377 | 5.868 | 5.875 |
| SIEM2 | 6.435 | 6.117 | 6.013 | 5.275 | 6.704 | 5.889 |
| BAYES | 6.629 | 5.995 | 6.322 | 5.349 | 6.530 | 5.779 |
| Skewness | | | | | | |
| $\ln(s^2)$ | -0.971 | -1.202 | -0.968 | -1.257 | -1.278 | -0.904 |
| QML | 0 | 0 | 0 | 0 | 0 | 0 |
| SIEM1 | -1.020 | -1.234 | -1.070 | -1.264 | -1.108 | -1.106 |
| SIEM2 | -0.886 | -1.402 | -0.831 | -1.230 | -1.369 | -0.580 |
| BAYES | -0.865 | -1.177 | -0.800 | -1.252 | -1.273 | -0.907 |
| Kurtosis | | | | | | |
| $\ln(s^2)$ | 1.225 | 2.022 | 0.895 | 2.850 | 2.731 | 1.181 |
| QML | 0 | 0 | 0 | 0 | 0 | 0 |
| SIEM1 | 1.127 | 2.244 | 1.441 | 2.350 | 1.598 | 1.363 |
| SIEM2 | 0.649 | 2.583 | 1.786 | 2.457 | 2.959 | 0.039 |
| BAYES | 0.337 | 1.544 | -0.123 | 2.065 | 2.385 | 0.356 |
| First order autocorrelation | | | | | | |
| $\ln(s^2)$ | 0.184 | 0.080 | 0.172 | 0.078 | 0.074 | 0.132 |
| QML | 0.174 | 0.105 | 0.192 | 0.054 | 0.131 | 0.124 |
| SIEM1 | 0.192 | 0.105 | 0.184 | 0.091 | 0.102 | 0.135 |
| SIEM2 | 0.132 | 0.071 | 0.156 | 0.080 | 0.061 | 0.095 |
| BAYES | 0.164 | 0.101 | 0.189 | 0.099 | 0.077 | 0.125 |
| Second order autocorrelation | | | | | | |
| $\ln(s^2)$ | 0.142 | 0.098 | 0.170 | 0.071 | 0.081 | 0.096 |
| QML | 0.170 | 0.102 | 0.184 | 0.053 | 0.086 | 0.104 |
| SIEM1 | 0.169 | 0.098 | 0.176 | 0.084 | 0.060 | 0.104 |
| SIEM2 | 0.129 | 0.069 | 0.151 | 0.077 | 0.058 | 0.088 |
| BAYES | 0.160 | 0.095 | 0.182 | 0.091 | 0.071 | 0.108 |

Notes: QML, SIEM1, SIEM2, BAYES represent the Quasi Maximum Likelihood, the Simulated EM with fixed and free mixture parameters and the Bayesian methods for estimating the univariate SV model, respectively. See text for details on computing moments.

two parameters ϕ and σ_η^2 , even asymptotically. The first order autocorrelation of y_t is

$$\text{AR}(1) = \text{corr}(y_t, y_{t-1}) = \frac{\phi\sigma_h^2}{\sigma_h^2 + \omega^2}$$

Keeping the variance of y_t fixed, $\text{Var}(y) = \sigma_h^2 + \omega^2$, we obtain

$$\text{AR}(1) = \phi(1 - \omega^2/\text{Var}(y)) \quad (3.10)$$

From (3.10) we see that increasing ω^2 must lead to a larger value of ϕ to keep the sample first order autocorrelation $\text{AR}(1)$ constant. This is exactly what we find in the estimates in Table 3.3, and most clearly for the pound/yen rate. In general ω^2 will be higher than $\pi^2/2$ if the distribution of the underlying exchange innovation ϵ_t has fat tails. In this way the distribution of ϵ_t and the estimates of the volatility persistence are connected. QML and SIEM1 can be seen as a restriction on one of the parameters in an ARMA(1,1) process. The modified QML estimator does yield higher estimates of the volatility persistence ϕ , but generally not as high as SIEM2. The implicit outlier correction of SIEM2, due to the high estimated kurtosis, must explain the remaining (small) differences.

An ARMA(1,1) process is fully identified by the variance and the first two autocorrelations. Ignoring all distributional identifying restrictions one could estimate the parameters of the stochastic volatility model from these three moments. However, the second order autocorrelation is larger than the first order autocorrelation, so that the implied estimate $\phi = \text{AR}(2)/\text{AR}(1)$ is larger than one, implying an explosive volatility process. This means that a GMM estimator with more moments is required. Given the number of observations the GMM criterion function will reject the overidentifying moment conditions, implying misspecification of the simple first order stochastic volatility model. The same conclusion also follows from comparing the sample autocorrelations and the implied autocorrelations in Table 3.7. None of the estimators provides a uniformly good fit for all six exchange rates. Since the model does not seem to fit the lower order moments care must be taken in a GMM estimation exercise. Both the choice and the number of moments seem to be important, as was also shown in Andersen and Sørensen (1994).

Instead of a GMM-based misspecification analysis, we computed a set of Lagrange Multiplier tests, taking the free mixture SIEM2 model as the null hypothesis. The tests take the form of adding an explanatory variable in the volatility equation

$$h_t = \phi h_{t-1} + \gamma h_{t-1}^* + \eta_t,$$

where h_{t-1}^* is the additional variable. Candidate additional variables are a second lag of the own volatility (h_{t-2}) of a currency in order to test for dynamic misspecification, and the lagged volatility of another exchange rate in order to test for spillover effects. We

TABLE 3.8: Diagnostic LM tests

| | JP/US | GE/US | UK/US | GE/JP | UK/JP | GE/UK |
|-------------------------------|-------|-------|-------|-------|-------|-------|
| Cross volatility tests | | | | | | |
| JP/US | — | 6.34 | 4.49 | 4.49 | 7.32 | 0.70 |
| GE/US | 0.88 | — | 9.57 | 4.89 | 1.29 | 0.68 |
| UK/US | 3.39 | 5.78 | — | 3.07 | 9.27 | 1.35 |
| GE/JP | 0.61 | 8.85 | 1.92 | — | 7.79 | 3.84 |
| UK/JP | 4.56 | 1.96 | 4.41 | 16.52 | — | 8.24 |
| UK/GE | 0.17 | 0.17 | 1.04 | 1.74 | 2.92 | — |
| Higher order dynamics | | | | | | |
| h_{t-2} | 63.11 | 68.78 | 98.39 | 82.35 | 73.45 | 85.56 |

Notes: The LM statistics test for inclusion of a cross-volatility h_{t-1}^* , or a further lag of own volatility in the volatility equation (h_{t-2}). The scores are all computed by simulation under the null hypothesis of the free mixture model.

computed LM statistics using the simulated states from the SIEM2 model and numerical differentiation of the likelihood. The scores were computed from a simulation with the multimove Gibbs sampler under the null hypothesis. The statistics are asymptotically distributed as a chi-square with one degree of freedom.

Table 3.8 shows that spillovers seem to be present in a number of cases, especially for the pound/yen rate. Note also that among exchange rates that do not have a certain currency in common (the anti-diagonal in the table) no evidence for spillovers exist. These results direct the interest towards a multivariate model of exchange rates. This idea is pursued in the next chapter. The validity of the dynamic specification of the volatility process is rejected overwhelmingly for all series.

3.4 Volatility Estimates

Even though the first order SV model appears misspecified, it has a strong theoretical appeal in finance applications, and remains an improvement on the assumption of constant volatility. It is therefore of interest to investigate the estimated volatilities of the different estimators.

The volatilities can be estimated in three different ways. Conditional on the parameters

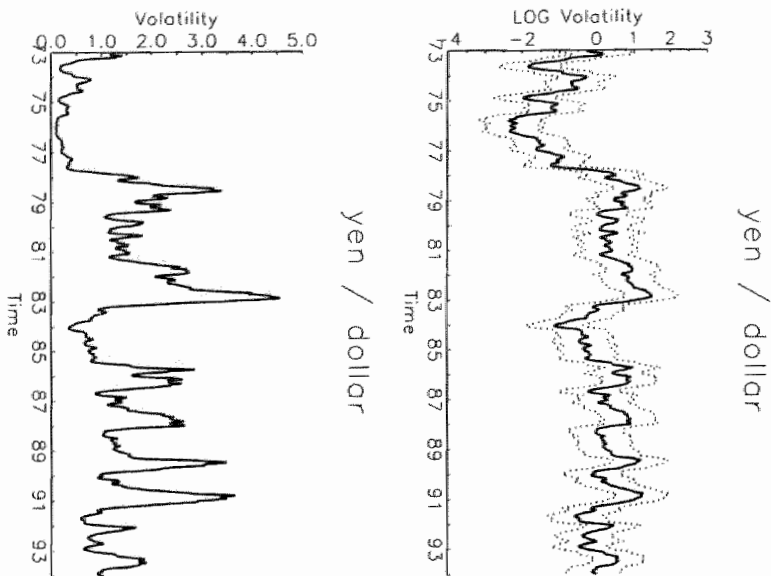


FIGURE 3.6: Estimated Volatilities: yen/dollar

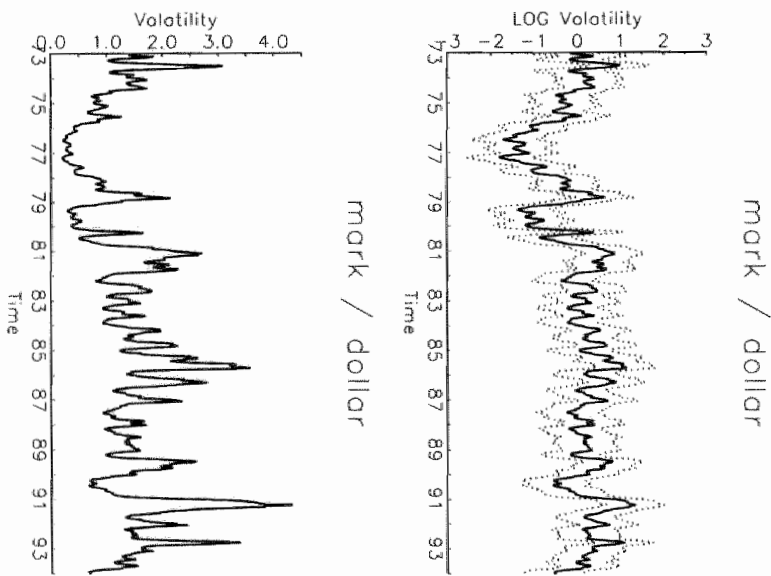


FIGURE 3.7: Estimated Volatilities: mark/dollar

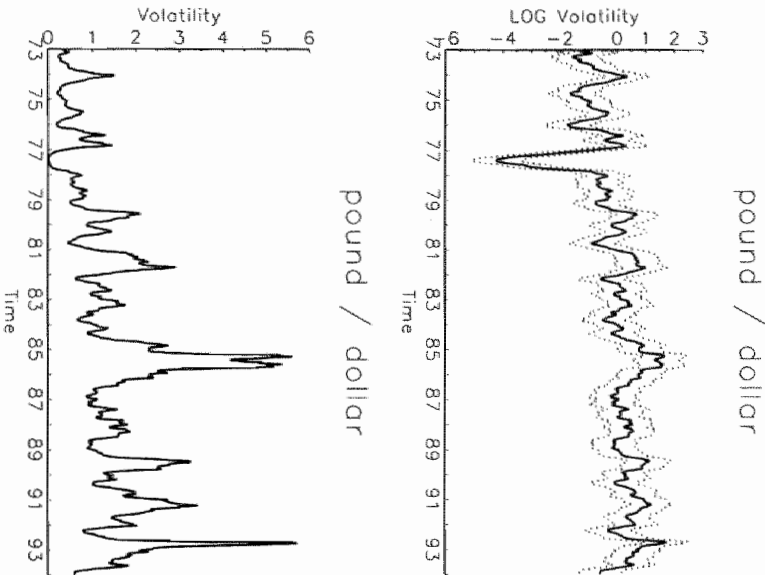


FIGURE 3.8: Estimated Volatilities: pound/dollar

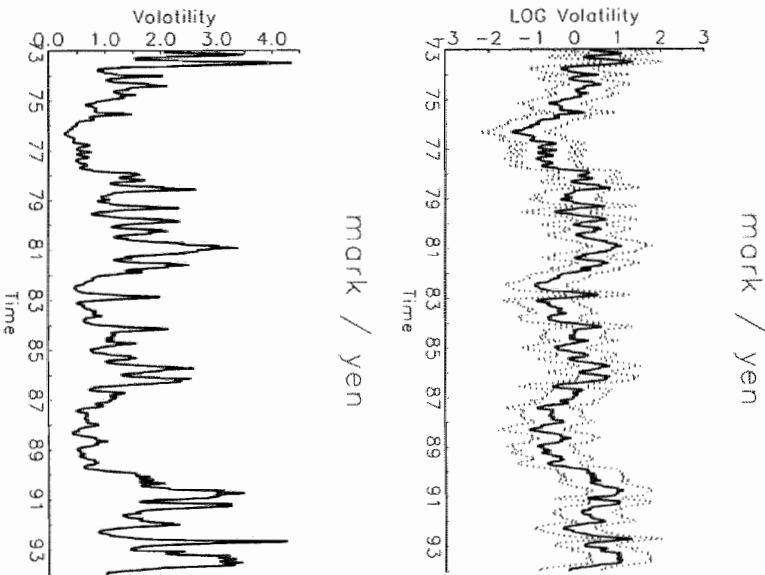


FIGURE 3.9: Estimated Volatilities: mark/yen

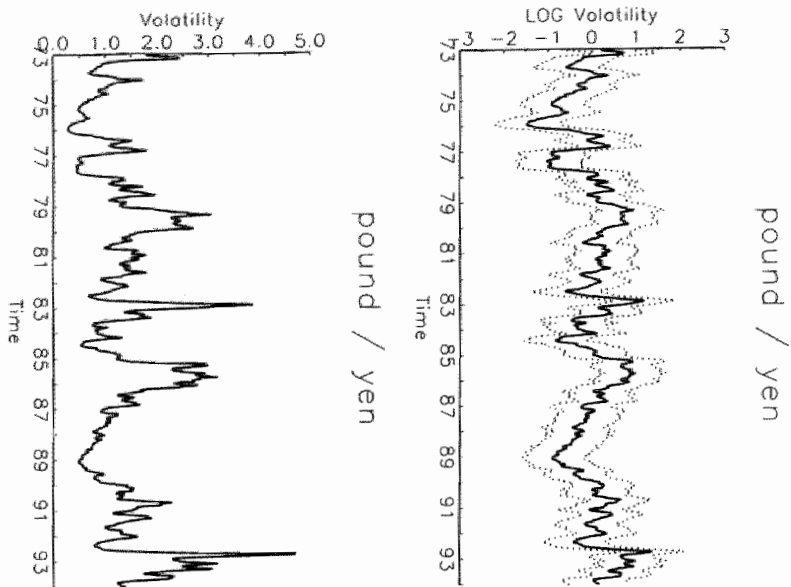


FIGURE 3.10: Estimated Volatilities: pound/yen

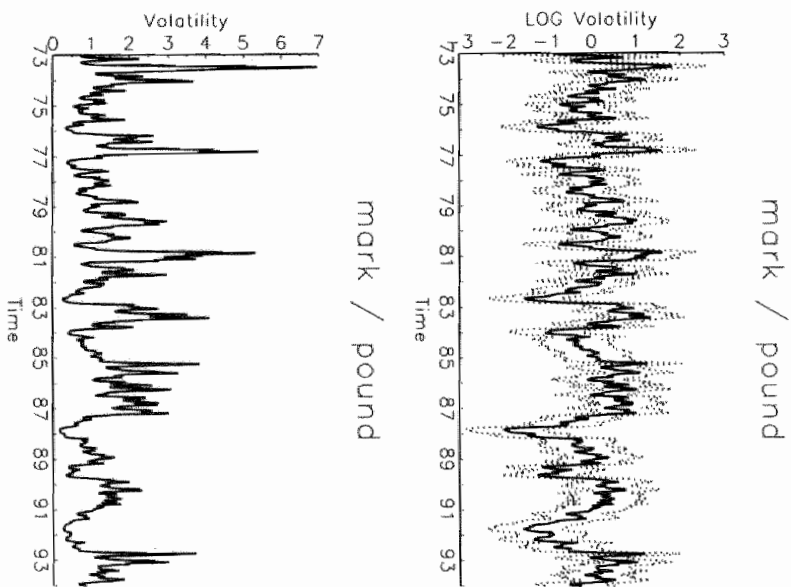


FIGURE 3.11: Estimated Volatilities: mark/pound

the Kalman smoother produces estimates under the assumption that the measurement error ξ_t is normally distributed. An alternative is to run the Gibbs sampler conditional on the parameters. The Gibbs sampler alternates between simulating from the states using $f(H_T|Z_T, Y_T; \theta)$ and the mixture indicators $f(Z_T|H_t, Y_T; \theta)$ as in the Simulated EM algorithm. The third volatility estimate is the Bayesian posterior mean of H_T , which is obtained as part of the full Gibbs sampler in the BAYES estimator. The first two volatility estimates are conditional on the parameters, while the Bayesian estimate incorporates parameter uncertainty. The simulation estimators also provide estimates of the actual variances $\exp(h_t)$. Because the unconditional mean β is not identified in the BAYES and SIEM2 models, we subtract the sample mean of the smoothed h_t . Similarly we also correct the sample mean of the estimated $\exp(h_t)$ by multiplying with a constant such that it equals the unconditional sample variance of the exchange rates $\Delta \ln S_t$.

In Table 3.9 we present some summary statistics of the estimated log-volatilities. The total variance σ_h^2 can approximately be split in two components,

$$\sigma_h^2 = \text{Var}(\hat{h}_t) + \bar{P}^2,$$

where \bar{P}^2 is the sample average of the variances of the estimation errors $(h_t - \hat{h}_t)$, which is obtained for each period t as the variance of the simulated $h_t^{(i)}$. The implicit assumption is that both components are orthogonal. All three components are reported in Table 3.9. The lowest relative estimation error variance \bar{P}^2/σ_h^2 is obtained for the dollar/pound rate, where the sample variance of \hat{h}_t accounts for 90% of the total variation of the volatility. The highest error ratios are for the cross rates, where the sample variance of \hat{h}_t is only between 60% and 70% of the total variance. The estimation error, conditional on the parameters, is also big in absolute value. Noting that the numbers refer to log-volatility we find that an average 95% confidence interval for a weekly volatility ranges roughly from a half to two times the estimated volatility. The BAYES estimator has even larger standard errors as it also incorporates the parameter uncertainty. For any financial application this is a very wide range. These numbers are further evidence for the need to improve the specification of SV models. KS do a first step in this direction by decomposing the volatility into a permanent and a transitory component as in Engle and Lee (1993). Other specifications are given in Jacquier, Polson and Rossi (1995), who study leptokurtic and skewed measurement error distributions, and correlated measurement and state error distributions in a univariate SV model. We leave these issues for further research.

The lower part of the table considers the differences between the SIEM2 simulation smoother, and alternative smoothing algorithms. Given that all series have mean zero by construction, we use the standard deviation of the time series of the difference between two series as a distance measure. These standard errors are large. Even if the same para-

TABLE 3.9: Summary Statistics of Volatility Estimates

| | JP/US | GE/US | UK/US | GE/JP | UK/JP | GE/UK |
|--|-------|-------|-------|-------|-------|-------|
| SIEM2 simulation estimates: \hat{h}_t | | | | | | |
| $SD(\hat{h}_t)$ | 0.837 | 0.576 | 0.938 | 0.540 | 0.508 | 0.586 |
| $\hat{\sigma}_h$ | 0.932 | 0.666 | 0.985 | 0.667 | 0.655 | 0.775 |
| Standard errors of \hat{h}_t: \bar{P}_t | | | | | | |
| SIEM2 | 0.421 | 0.343 | 0.459 | 0.394 | 0.408 | 0.487 |
| BAYES | 0.549 | 0.485 | 0.578 | 0.497 | 0.497 | 0.626 |
| SIEM2 (simulated) minus SIEM2 (Kalman) | | | | | | |
| std. dev. | 0.244 | 0.266 | 0.244 | 0.241 | 0.505 | 0.240 |
| AR(1) | 0.082 | 0.121 | 0.041 | 0.067 | 0.647 | 0.063 |
| SIEM2 (simulated) minus QML (Kalman) | | | | | | |
| std. dev. | 0.248 | 0.267 | 0.270 | 0.385 | 0.330 | 0.264 |
| AR(1) | 0.156 | 0.213 | 0.195 | 0.512 | 0.321 | 0.254 |
| SIEM2 (Kalman) minus QML (Kalman) | | | | | | |
| std. dev. | 0.082 | 0.111 | 0.121 | 0.250 | 0.450 | 0.159 |
| AR(1) | 0.018 | 0.004 | 0.055 | 0.038 | 0.759 | 0.061 |
| SIEM2 (Kalman) minus BAYES | | | | | | |
| std. dev. | 0.091 | 0.145 | 0.103 | 0.106 | 0.082 | 0.166 |
| AR(1) | 0.110 | 0.379 | 0.075 | 0.230 | 0.136 | 0.445 |

Notes: $SD(\hat{h}_t)$ is the sample standard deviation of the time series \hat{h}_t of estimated log-volatilities, using the multimove Gibbs sampler as the smoothing algorithm; σ_h is the estimated standard deviation of h_t using the parameter estimates in table 3.3; \bar{P}_t is the square root of the sample mean of the estimated error variances of $h_t - \hat{h}_t$. The lower part of the table provides information on the time series of differences between the SIEM2 simulated states and estimates obtained by the standard Kalman smoother. AR(1) is the first order autocorrelation of the series of deviations.

meters are used, the Kalman smoother produces a series that deviates strongly from the simulation smoother. The differences between the BAYES estimator and the SIEM2 simulation smoother are much smaller. There are also much smaller differences between the two series produced by the Kalman smoother. From this we conclude that the issue of the right smoother is much more important than the parameter estimation method in this SV model.

Table 3.9 also shows that the dollar/yen and dollar/pound exchange rates exhibit much wider fluctuations in the volatility than the other four series. The fluctuations are shown in

figures 3.6 to 3.11, which present the smoothed log volatilities h_t of the BAYES estimator in the top panel, and the actual volatilities $\exp(h_t)$ (scaled to match the unconditional sample variance) in the lower panel. The smoothed volatilities in the lower panel have been obtained as the average of the simulated $\exp(h_t^{(i)})$ over a run of the Gibbs sampler. The dollar/pound volatility has a large dip in 1977, when the exchange rate was very stable. The peaks of the dollar/pound volatility in 1985 and 1993 appear very sharply in the lower panel. The high sterling volatility starting in the fall of 1992 is also visible in the pound/mark and pound/yen rates, identifying this episode as sterling volatility.

3.5 Conclusions

In this chapter we have discussed some powerful algorithms to estimate the hyperparameters and latent variables of a general non-Gaussian state-space model. Two methods were analyzed: the Bayesian Gibbs sampler and the Simulated EM algorithm. We empirically studied the performance of the first order lognormal SV model using a dataset of weekly exchange rates over the last twenty years. The model has been estimated by estimation methods that differ in the specification of the distribution of the standardized exchange rate innovations. The main finding is that, for the number of observations we use, the point estimates of the persistence of the volatility increase if the error distribution is explicitly modelled through a mixture of normals. The estimation method (Simulated EM or Bayesian Gibbs sampling) does not seem to make any difference in this respect.

The other results of this chapter pertain to the estimation of the time series of volatilities. First we find that different smoothing algorithms produce very different estimates, even if the parameters of the underlying process are the same. Again the differences arise from explicit consideration of the measurement error density in the state-space model for the log volatility. The most disturbing finding is that even the most efficient simulation smoothers produce very large standard errors for the volatility estimates.

Finally, diagnostic tests reveal several forms of misspecification of the standard stochastic volatility model. In particular, multivariate specifications and higher order dynamics seem to be attractive empirical extensions.

Appendix 3.A: Bayesian Simulation Algorithm

The Bayesian Gibbs sampler for the stochastic volatility model has the following steps:

1. Let $\mu_t = \mu(z_t)$ and $\omega_t^2 = \omega^2(z_t)$, $t = 1, \dots, T$ be the mean and variance associated with the current draw for the index of the element of the mixture distribution. Conditional on z_t and the parameters σ_η , μ , and ω the Kalman filter produces an estimate of the filtered states $\hat{h}_t = E[h_t|Y_t]$. The recursion is

$$\begin{aligned}\hat{h}_t &= \phi \hat{h}_{t-1} + \frac{\hat{p}_t^2}{\omega_t^2} (y_t - \phi \hat{h}_{t-1} - \mu_t) \\ \hat{p}_t^2 &= \frac{\omega_t^2 (\phi^2 \hat{p}_{t-1}^2 + \sigma_\eta^2)}{\phi^2 \hat{p}_{t-1}^2 + \sigma_\eta^2 + \omega_t^2},\end{aligned}$$

where \hat{p}_t^2 is the conditional variance of \hat{h}_t . The initial conditions of the filter are:

$$\begin{aligned}\hat{h}_1 &= y_1 - \mu_1 \\ \hat{p}_1^2 &= \omega_1^2\end{aligned}$$

2. Given the filtered estimates \hat{h}_T and \hat{p}_T , the final state h_T^s is generated as a draw from $N(\hat{h}_T, \hat{p}_T^2)$. A backward recursion produces the new time series of simulated states. The recursion is taken from Shephard (1994). For $t = T-1, \dots, 1$, generate

$$h_t^s \sim N \left[\hat{h}_t + \frac{\phi \hat{p}_t^2}{\phi^2 \hat{p}_t^2 + \sigma_\eta^2} (h_{t+1}^s - \phi \hat{h}_t), \frac{\sigma_\eta^2}{\phi^2 \hat{p}_t^2 + \sigma_\eta^2} \right]$$

3. Given a simulation for the time series $H_T = (h_1^s, \dots, h_T^s)'$ a new draw for the autocorrelation parameter ϕ is generated from the conditional posterior of ϕ , which is truncated normal with mean and variance parameters $\hat{\phi}$ and \hat{S}_ϕ that are obtained by the OLS regression of h_t^s on h_{t-1}^s . Since the prior of ϕ is truncated to the interval $(-1, 1)$ a draw from the conditional normal is only accepted if it falls in this region.
4. Given ϕ and H_T a new value for σ_η is drawn from the inverted Gamma distribution with degrees of freedom $(T-1 + d_0)$ and sum of squares $(s_0 + \text{SSR})$, where SSR is the sum of squared residuals of the regression of h_t^s on h_{t-1}^s in 3, and where d_0 and s_0 are the prior degrees of freedom and prior sum of squares in the prior on σ_η .
5. Conditional on the data Y_T and the states H_T , the innovations ξ_t are observable as $y_t - h_t^s$. Given the regime indicators z_t , the posterior of the means μ_i ($i = 1, 2, 3$) of the mixtures is implicitly given by the regression models

$$\xi(i) = \mu_i t + v(i),$$

where $\xi(i)$ denotes a vector of length T_i with elements corresponding to $z_t = i$, and where ι is a T_i vector of ones. The error term in this model has mean zero and variance ω_i^2 . Since the prior on μ_i is $N(m_i, V_{0i}^2)$, the conditional posterior of μ_i is also normal with mean and variance given by

$$\hat{\mu}_i = \frac{T_i V_{0i}^2 \bar{\xi}_i + \omega_i^2 m_i}{T_i V_{0i}^2 + \omega_i^2}, \quad \text{and} \quad \left(\frac{1}{V_{0i}^2} + \frac{T_i}{\omega_i^2} \right)^{-1},$$

respectively, with $\bar{\xi}_i$ the sample mean of $\xi(i)$. These densities are used to draw new values for the μ_i . The unconditional mean β of ξ follows as $\beta = \sum_{i=1}^3 p_i \mu_i$.

6. Analogolously to drawing σ_η we draw ω_i from an inverted Gamma distribution, in this case with degrees of freedom parameter $(T_i - 1 + d_0)$ and sum of squares $(\omega_{0i}^2 + \text{SSR}_i)$, with SSR_i the sum of squared residuals from the regression in 5.
7. Draw new mixture indicators z_t ($t = 1, \dots, T$) as in the SIEM algorithms described in the text.

This completes a single iteration of the Gibbs sampler in a Bayesian parameter estimation exercise.

Appendix 3.B: Moments

The first four non-central moments of a random variable z distributed as a mixture of normals with means and variances μ_i and ω_i^2 , ($i=1, \dots, M$) are given by

$$E[z^j] = \sum_{i=1}^M p_i (j^{\text{th}} \text{ moment of } N(\mu_i, \omega_i^2)),$$

where p_i is the weight of the i^{th} normal density in the mixture. This results in the following four moments

$$\begin{aligned} m_1 = E[z_t] &= \sum_{i=1}^M p_i \mu_i, \\ m_2 = E[z_t^2] &= \sum_{i=1}^M p_i (\omega_i^2 + \mu_i^2), \\ m_3 = E[z_t^3] &= \sum_{i=1}^M p_i (3\mu_i \omega_i^2 + \mu_i^3), \\ m_4 = E[z_t^4] &= \sum_{i=1}^M p_i (3\omega_i^4 + 6\mu_i^2 \omega_i^2 + \mu_i^4), \end{aligned} \tag{3.11}$$

The non-central moments of the independent variable of the simple univariate stochastic volatility models can be found using the moments above. Consider the model

$$\begin{aligned} y_t &= h_t + \xi_t, \\ h_t &= \phi h_{t-1} + \eta_t, \quad \eta_t \sim N((0, \sigma_\eta^2), \end{aligned}$$

where ξ_t is independent of η_t at all leads and lags. The first four unconditional moments of y_t and the first two time series moments are given by

$$\begin{aligned} E[y_t] &= m_1, \\ E[y_t^2] &= \sigma_h^2 + m_2, \\ E[y_t^3] &= 3m_1 \sigma_h^2 + m_3, \\ E[y_t^4] &= 3\sigma_h^4 + 6m_2 \sigma_h^2 + m_4, \\ E[y_t y_{t-1}] &= m_1^2 + \phi \sigma_h^2, \\ E[y_t y_{t-2}] &= m_1^2 + \phi^2 \sigma_h^2, \end{aligned}$$

where $\sigma_h^2 = \sigma_\eta^2 / (1 - \phi^2)$. In case the measurement equation contains the constant term β one should replace all μ_i in (3.11) by $\mu_i + \beta$. The central moments can be computed from these moments in the usual way.

To obtain the moments of $\epsilon_t = \pm \epsilon_t^{1/2}$ we assume that the distribution of ϵ_t is symmetric so that all odd moments are zero. The even moments are found analogous to (3.11) by

using the properties of the individual elements in the mixture:

$$\begin{aligned} \mathbb{E}[\epsilon_t^2] &= \mathbb{E}[e^{\epsilon_t}] = \sum_{i=1}^M p_i \exp(\mu_i + \omega_i^2/2), \\ \mathbb{E}[\epsilon_t^4] &= \mathbb{E}[e^{2\epsilon_t}] = \sum_{i=1}^M p_i \exp(2\mu_i + 2\omega_i^2). \end{aligned}$$

Chapter 4

Neglected Common Factors in Exchange Rate Volatility

The empirical literature on floating exchange rates has largely concentrated on the behavior of the dollar against the major other currencies like the German mark, Japanese yen, and British pound. The cross rates have attracted much less attention. One of the stylized facts from the literature is that the time series of the logarithm of the dollar exchange rates is close to a random walk, meaning that almost all changes in the dollar can be interpreted as unpredictable news. Another stylized fact is the conditional heteroskedasticity in all time series of exchange rate changes. In this chapter we develop a new model for exchange rate volatility that simultaneously describes the volatility in all possible bilateral exchange rates between the major currencies and does not depend on a particular numeraire currency.

In the empirical model it is assumed that the change in the logarithm of any bilateral exchange rate is the difference of two country specific news terms. This model implies that the first differences of the exchange rates are positively correlated. For example, if the dollar rises or falls with respect to both the mark, the yen, as well as the pound, it is likely that there has been some important news about the U.S. economy. In general, the higher the correlation between the exchange rates of the dollar/yen and dollar/mark, the larger is the U.S. news component in daily or weekly exchange rate changes. The separate components are identifiable by exploiting the triangular identity which states that the difference between the log of the dollar/mark and the dollar/yen exchange rate yields the log of the yen/mark exchange rate. This enables us to perform a variance decomposition of exchange rate changes.

Univariate models of conditional heteroskedasticity are abundant, but relatively few studies use a multivariate framework. The main obstacle here is the large number of parameters

involved in an unrestricted model for the time variation in volatility. For estimation a large number of usually ad hoc restrictions have to be imposed. There is still a quest for a convenient parameterization of multivariate volatility models, which can meet the empirical success of Bollerslev's (1986) univariate GARCH(1,1) model. Our decomposition of exchange rates in country specific news components provides an parsimonious parameterization of multivariate volatility dynamics.

The decomposition results are in principle applicable to various functional forms and stochastic specifications, e.g. Bollerslev's (1986) GARCH model or Nelson's (1991) EGARCH framework. However, in the empirical part of the paper we follow the ideas of the stochastic variance model of Harvey, Ruiz, and Shephard (1993). Their model also aims at a parsimonious parameterization, and also involves a factor structure, which makes their approach closely related to ours. One advantage of the stochastic volatility model is its flexibility in specifying the dynamics and in dealing with fat-tailed distributions. The requirement of numeraire invariance imposes some further structure on their model.

One motivation for the factor model that we adopt in the paper comes from the covariance structure that Domowitz and Hakkio (1985) used to derive their model of risk premia in the foreign exchange market. They showed that risk premia depend on the difference between the conditional variances of the country's money supplies, which are the only stochastic elements in their model. In our approach we make a factor structure assumption to identify these news components individually.

This chapter is organized as follows. In Section 4.1 we develop the variance decomposition and report empirical results for weekly exchange rate changes for the period 1973-1991 and several subperiods. Section 4.2 presents the generalization of the variance decomposition to models with time varying volatility. Section 4.3 reports empirical results for this model. In Section 4.4 we look at the asset pricing implications. Finally, section 4.5 concludes.

4.1 A Factor Structure of Exchange Rate News

Consider a system of $n+1$ currencies ($i = 0, \dots, n$), and express bilateral exchange rates with respect to the common numeraire currency 0. We assume that exchange rate changes are almost unpredictable and due to news. News in each country has two parts: a component related to worldwide shocks and a country specific component. This setup leads to the following model of exchange rate movements:

$$s_{i0} = u_i - u_0, \quad i = 1, \dots, n, \quad (4.1)$$

$$u_i = \sum_{k=1}^M \beta_{ik} v_k + e_i, \quad i = 0, \dots, n,$$

where s_{i0} is the change in the logarithm of the bilateral exchange rate of currency i in units of currency 0; u_i is the news originating from country i ; v_k is a worldwide common factor of news; β_{ik} is the sensitivity of news in country i with respect to worldwide shock v_k ; and e_i is the idiosyncratic news component. It is assumed that all factors have zero mean and are mutually uncorrelated. The variances of the common factors are normalized to one, while the variances of the country specific factors are $E(e_i^2) = \lambda_i^*$. The difference between Equation (4.1) and a standard linear factor model is that the outputs u_i are only observed through the exchange rates s_{i0} . Combining (4.1) and (4.2) gives

$$s_{i0} = \sum_{k=1}^M (\beta_{ik} - \beta_{0k}) v_k + e_i - e_0 = \sum_{k=0}^M \gamma_{ik} v_k + e_i, \quad (4.2)$$

where $\gamma_{i0} = -1$, $v_0 = e_0$, and $\gamma_{ik} = (\beta_{ik} - \beta_{0k})$ for $k \geq 1$. Representation (4.2) explains the term "neglected" in the title of this chapter. The first common factor in exchange rates is the numeraire specific news e_0 . The common factors v_1, \dots, v_M only affect exchange rates with loadings $(\beta_{ik} - \beta_{0k})$. Even if the common factors are important in the total currency news u_i , they might not have much effect on the exchange rate changes, when β_{ik} and β_{jk} are approximately equal.

To determine the number of common factors one would need a large number of currencies. In the empirical analysis we will limit ourselves to the four major international currencies. By concentrating on these four currencies we implicitly assume that $M = 4$. To estimate the model on just four currencies we will have to make some assumptions on the structure of the factor loadings β_{ik} . In particular we assume that (4.2) has the form

$$\begin{pmatrix} u_{US} \\ u_{JP} \\ u_{GE} \\ u_{UK} \\ \mathbf{u}_{rest} \end{pmatrix} = \begin{pmatrix} \beta_{01} & 0 & 0 & 0 \\ 0 & \beta_{12} & 0 & 0 \\ 0 & 0 & \beta_{23} & 0 \\ 0 & 0 & 0 & \beta_{34} \\ & & \mathbf{B}^* & \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{pmatrix} + \begin{pmatrix} e_{US} \\ e_{JP} \\ e_{GE} \\ e_{UK} \\ \mathbf{e}_{rest} \end{pmatrix}$$

where the subscripts US, JP, GE, UK denote the dollar, yen, mark and pound; \mathbf{u}_{rest} is the $(n-3)$ vector of news of all other currencies, and \mathbf{e}_{rest} the specific news; the matrix of factor loadings is assumed to be diagonal in the major currencies, but \mathbf{B}^* is unrestricted. This entails just 2 overidentifying restrictions on a general four factor model. Under these assumption the covariance matrix of the four major currencies with the dollar as numeraire takes the form

$$\Sigma = \begin{pmatrix} \lambda_0 + \lambda_1 & \lambda_0 & \lambda_0 \\ \lambda_0 & \lambda_0 + \lambda_2 & \lambda_0 \\ \lambda_0 & \lambda_0 & \lambda_0 + \lambda_3 \end{pmatrix}, \quad (4.3)$$

where we have defined $\lambda_i \equiv \lambda_i^* + \beta_{i,i+1}^2$. This model is equivalent to a zero factor model for just four currencies, and preserves the two overidentifying restrictions. The model implies that all covariances are equal and positive. For our weekly dataset of the four major currencies (dollar, yen, mark and pound) the observed sample covariances for the period 1973–1991 with the dollar as the numeraire are

$$\Sigma = \begin{pmatrix} 2.09 & 1.31 & 1.06 \\ 1.31 & 2.23 & 1.49 \\ 1.06 & 1.49 & 2.10 \end{pmatrix}.$$

At first sight this covariance matrix is remarkably close to that implied by the zero factor model (4.3) with all λ_i equal, so that the model merits a closer statistical investigation.

The theoretical literature on foreign exchange risk premia has mostly considered the case $M=0$. In Domowitz and Hakkio (1985) the e_i are uncorrelated unexpected shocks to each country's money supply. In a more general setting Hodrick (1989) develops an equilibrium model of the exchange rate which can also be written as the factor model with $M=0$, and where e_i is a linear combination of news of each country's money growth, real output growth, government expenditures, and volatility innovations. In Hodrick's model each of the currency specific factors are assumed (conditionally) heteroskedastic.

We will estimate the variance λ_i for several subsamples of weekly exchange rate data. Since normality is always strongly rejected for exchange rate changes (see, *e.g.*, Boothe and Glassman, 1987), we adopt a moment estimator that does not rely on normality. Let $y(t)$ be the $\frac{1}{2}n(n+1)$ vector containing the squares of all possible bilateral exchange rate changes between $n+1$ currencies, with $y_k(t) = s_{ij}^2(t)$ as the k th element of $y(t)$. Let \mathbf{Z} be the $(\frac{1}{2}n(n+1) \times (n+1))$ matrix, with rows containing all permutations of two ones and $(n-1)$ zeros in $(n+1)$ positions. The ones in the k th row of \mathbf{Z} are in the positions i and j and correspond to the squared bilateral exchange rate change $y_k(t) = s_{ij}^2(t)$. Since the variance of every bilateral exchange rate s_{ij} is modelled as the sum of λ_i and λ_j , we can formulate the linear model

$$y(t) = \mathbf{Z}\lambda + \nu(t), \quad (4.4)$$

where $\nu(t)$ is a vector of errors with mean zero. Omitting the time indices, the model for $n=3$ is written as

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} \lambda_0 \\ \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{pmatrix} + \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \\ \nu_4 \\ \nu_5 \\ \nu_6 \end{pmatrix}. \quad (4.5)$$

The parameter vector λ can be estimated consistently by OLS, pooling the time series and the cross section of six bilateral exchange rate changes. To construct an efficient estimator,

we can use the initial OLS estimates to form the $(\frac{1}{2}n(n+1) \times \frac{1}{2}n(n+1))$ weighting matrix $\tilde{\mathbf{D}}$ with typical elements $\tilde{d}_{ij,km}$ as

$$\tilde{d}_{ij,km} = \frac{1}{T} \sum_{t=1}^T (s_{ij}^2(t) - \tilde{\lambda}_i - \tilde{\lambda}_j) (s_{km}^2(t) - \tilde{\lambda}_k - \tilde{\lambda}_m).$$

Applying SUR we find an efficient moment estimator for λ as

$$\hat{\lambda} = (\mathbf{Z}'\tilde{\mathbf{D}}^{-1}\mathbf{Z})^{-1} (\mathbf{Z}'\tilde{\mathbf{D}}^{-1}\bar{\mathbf{y}}) \quad (4.6)$$

where $\bar{\mathbf{y}}$ is the (6×1) vector containing the second moments of the six bilateral exchange rates. The weighting implied by this estimator is inversely related to the fourth moment of the exchange rate changes. In computing the estimates we maintain the assumption that the variances exist. Standard errors are computed by the usual formula, $\mathbf{V}(\hat{\lambda}) = (\mathbf{Z}'\tilde{\mathbf{D}}^{-1}\mathbf{Z})^{-1}/T$.

Given estimates of all the individual variances we can estimate time series of the country specific news. The observed n exchange rates impose n exact conditions on the $n+1$ individual uncorrelated news components. Formally, we want the conditional mean $E[e(t)|s_0(t)]$, where $e(t)$ is the $n+1$ news vector at time t , and $s_0(t)$ is the n -vector of exchange rate changes with numeraire currency 0. The GLS estimator for this conditional mean is given by

$$e(t) = \Lambda \mathbf{P}' (\mathbf{P} \Lambda \mathbf{P}')^{-1} s_0(t),$$

where $\mathbf{P} = (-\iota|\mathbf{I})$ is an $(n \times (n+1))$ matrix, and \mathbf{I} the n -dimensional identity matrix, and Λ is an $((n+1) \times (n+1))$ diagonal matrix with the variances of the specific factors on its diagonal. The specific elements of $\hat{e}(t)$ can also be written in the more explicit form

$$\hat{e}(t) = \frac{\sum_{j=0}^n \lambda_j^{-1} s_{ij}(t)}{\sum_{j=0}^n \lambda_j^{-1}}$$

The time series $\hat{e}_i(t)$ can be interpreted as the changes in the effective exchange rate of currency i .

Our dataset consists of the bilateral exchange rates among the dollar, yen, mark and pound for the period January 1973 to June 1991 and several subsamples. The data are weekly Wednesday closing prices at the London market taken from DATASTREAM. If the market was closed on some Wednesday, we choose the Tuesday price. If Tuesday was a holiday too, we took the Thursday price. In the empirical analysis we take all exchange rate changes in deviation of their sample mean.

The variance decomposition results are summarized in tables 4.1 and 4.2. Table 4.1 reports the estimates of the variances of the currency specific news components. The variances

TABLE 4.1: Variance Decomposition of Exchange Rate News

$$\text{Var}(s_{ij}) = \lambda_i + \lambda_j$$

| Period | λ_{US} | λ_{JP} | λ_{GE} | λ_{UK} | EQUAL | FIT |
|-------------------------------------|----------------|----------------|----------------|----------------|-------|------|
| Jan 73 - Jun 91 (Full sample) | 1.15 (0.10) | 0.93 (0.09) | 0.67 (0.07) | 0.70 (0.08) | 18.9 | 43.6 |
| Jan 73 - Dec 76 | 0.45 (0.11) | 0.75 (0.19) | 1.08 (0.23) | 0.74 (0.19) | 8.17 | 7.05 |
| Jan 77 - Dec 80 | 0.76 (0.12) | 1.40 (0.21) | 0.54 (0.11) | 0.54 (0.11) | 17.6 | 6.66 |
| Jan 81 - Sep 85 | 1.17 (0.13) | 0.69 (0.14) | 0.62 (0.13) | 0.84 (0.18) | 10.9 | 26.2 |
| Sep 85 - Feb 87 (Plaza - Louvre) | 1.32 (0.59) | 0.43 (0.37) | 0.73 (0.20) | 0.89 (0.32) | 1.84 | 12.4 |
| Feb 87 - Jun 91 | 1.77 (0.20) | 0.94 (0.14) | 0.40 (0.07) | 0.44 (0.09) | 50.3 | 5.86 |

Notes: λ_i Denotes the exchange rate variance related to country i . The estimates are from the second round of the moment estimator (4.6). Standard errors in parentheses. The column *EQUAL* is a Wald test for equality of all 4 λ_i 's, asymptotically distributed as $\chi^2(3)$. The column *FIT* is a Wald test for the overidentifying moment restrictions implied by the factor structure, asymptotically distributed as $\chi^2(2)$. Exchange rate changes are measured in percentage per week and are corrected with the sample mean.

of the specific currencies are significantly different from each other and are also different between subperiods. Over the full 1973-1991 sample the U.S. variance has contributed most to the volatility of the exchange rate system. The ranking of the variances is: $\lambda_{US} > \lambda_{JP} > \lambda_{UK} > \lambda_{GE}$, although the absolute differences between the currency specific variances are not very large. The dominance of the U.S. variance is especially due to the later part of the sample period. Over the last five years (87-91) the U.S. variance is four times as big as the volatility originating in either the U.K. or Germany, while in the first years of the floating exchange rate period (73-76) most of the variance was due to events in West Germany. For the period 77-80 Japan had by far the largest variance.

The fourth moments in Table 4.2 generally imply excess kurtosis and rejection of normality, as is common for exchange rate changes. The period between September 1985 and February 1987 (Plaza - Louvre) has been the most volatile episode of the last 20 years. The estimates are considerably above their full sample averages. At the same time the three dollar

TABLE 4.2: Sample and Fitted Moments of s_{ij}

| | JP/US | GE/US | UK/US | GE/JP | UK/JP | UK/GE |
|---|-------|-------|-------|-------|-------|-------|
| January 1973 - June 1991 (Full sample) | | | | | | |
| Σ_{ij} | 2.09 | 2.23 | 2.10 | 1.70 | 2.07 | 1.35 |
| $\lambda_i + \lambda_j$ | 2.08 | 1.83 | 1.86 | 1.60 | 1.63 | 1.37 |
| $K(s_{ij})$ | 33.4 | 17.9 | 20.5 | 11.3 | 22.5 | 9.06 |
| January 1973 - December 1976 | | | | | | |
| Σ_{ij} | 1.39 | 1.88 | 1.17 | 1.81 | 1.70 | 1.86 |
| $\lambda_i + \lambda_j$ | 1.20 | 1.53 | 1.19 | 1.83 | 1.49 | 1.82 |
| $K(s_{ij})$ | 95.6 | 21.5 | 8.87 | 20.6 | 30.1 | 18.0 |
| January 1977 - December 1980 | | | | | | |
| Σ_{ij} | 2.31 | 1.44 | 1.21 | 1.96 | 2.29 | 1.13 |
| $\lambda_i + \lambda_j$ | 2.16 | 1.29 | 1.30 | 1.94 | 1.94 | 1.07 |
| $K(s_{ij})$ | 14.2 | 6.55 | 5.13 | 10.2 | 16.1 | 4.38 |
| January 1981 - September 1985 | | | | | | |
| Σ_{ij} | 1.86 | 2.61 | 2.99 | 1.54 | 2.41 | 1.41 |
| $\lambda_i + \lambda_j$ | 1.86 | 1.78 | 2.00 | 1.32 | 1.53 | 1.46 |
| $K(s_{ij})$ | 10.9 | 16.5 | 38.3 | 8.14 | 23.4 | 9.31 |
| September 1985 - February 1987 (Plaza - Louvre) | | | | | | |
| Σ_{ij} | 2.92 | 3.19 | 2.72 | 1.87 | 3.00 | 1.74 |
| $\lambda_i + \lambda_j$ | 1.75 | 2.04 | 2.21 | 1.15 | 1.32 | 1.62 |
| $K(s_{ij})$ | 33.3 | 45.6 | 47.7 | 11.7 | 44.0 | 7.18 |
| February 1987 - June 1991 | | | | | | |
| Σ_{ij} | 2.48 | 2.44 | 2.52 | 1.50 | 1.54 | 0.81 |
| $\lambda_i + \lambda_j$ | 2.71 | 2.18 | 2.22 | 1.34 | 1.38 | 0.84 |
| $K(s_{ij})$ | 16.2 | 12.5 | 15.5 | 6.14 | 9.71 | 2.34 |

Notes: Σ_{ij} are the sample variances of exchange rate changes s_{ij} ; $\lambda_i + \lambda_j$ are the fitted second moments using the optimal moment estimator; $K(s_{ij})$ is the centered fourth moment required in the weighting matrix D, i.e. $E[(s_{ij}^2 - Es_{ij})^2]$. Exchange rate changes s_{ij} are measured in percentage per week and are corrected with the sample mean.

exchange rates and the yen/pound rate have much higher fourth moments than in other subperiods.

Returning to the results in Table 4.1 we find that the two overidentifying conditions implied by the factor structure (4.3) are strongly rejected for the full sample and also for the two subperiods between 1981 and February 1987. The test statistics do, however, not take

TABLE 4.3: Diagnostics of News Components

| | e_{US} | e_{JP} | e_{GE} | e_{UK} |
|-----------------|----------|----------|----------|----------|
| Skewness | 0.52 | -0.88 | -0.61 | 0.64 |
| Kurtosis | 2.36 | 4.24 | 2.67 | 3.33 |
| Normality | 268.4* | 849.1* | 347.6* | 511.8 |
| Autocorrelation | 0.057 | 0.031 | 0.100 | 0.066 |
| Ljung-Box (30) | 36.04 | 37.74 | 51.65* | 40.30 |
| ARCH(1) | 4.66* | 7.32* | 8.69* | 17.10* |
| ARCH(10) | 47.52* | 43.85* | 86.09* | 37.71* |
| Cross ARCH | 2.68 | 37.15* | 0.59 | 20.79* |
| Causality | 0.90 | 16.62* | 0.20 | 8.77* |

Notes: *Skewness* is the scaled third moment of the news series, $\sum e^3 / (\sum e^2)^{3/2}$; *Kurtosis* is the excess kurtosis ($\sum e^4 / (\sum e^2)^2 - 3$); *Normality* is the Jarque-Bera test for normality, distributed as $\chi^2(2)$; *Autocorrelation* is the first order autocorrelation; *Ljung-Box (30)* is a test for autocorrelation up to order 30, distributed as $\chi^2(30)$; *ARCH(1)* is the LM test for first order ARCH, distributed as $\chi^2(1)$; *ARCH(10)* is the LM test for 10th order ARCH ($\chi^2(10)$); *Cross ARCH* is TR^2 of the regression of $\hat{e}_i^2(t)$ on a constant and all four squared news series with one lag ($\chi^2(4)$); *Causality* is the F-statistic for the significance of the cross squared news series in the last cross ARCH regression. An asterisk (*) denotes significance at the 5% level.

into account any further heteroskedasticity within the subperiods. As is also suggested by Table 4.2 the model appears to fit well in the seventies and again after 1987.

Table 4.3 reports a set of diagnostics of the extracted news components using the full sample parameter estimates. The non-normality is not confined to one particular currency, but appears in all four series. There is hardly any autocorrelation in the news series, except for a slightly significant Ljung-Box statistic for the Japanese news series. In contrast, there is strong evidence of ARCH in all four components, indicating that the ARCH behaviour is not special to the U.S. dollar. More interesting is the finding that there are strong heteroskedasticity spillovers. The Granger causality tests indicate that the Japanese and British squared news components are predictable by the other countries' lagged squared news.

4.2 Multivariate Time Varying Volatility

4.2.1 Specification Problems

A multitude of specifications exist for modelling exchange rate volatility in univariate models. Most of these models are a variant of the ARCH model developed by Engle (1982). For exchange rates the empirical evidence favours a specification with fat-tailed errors even after correcting for the conditional heteroskedasticity (Engle and Bollerslev, 1986; Baillie and Bollerslev, 1989, and Hsieh, 1989). The main problem in specifying a multivariate model is the number of parameters that is of the order n^4 when the univariate models are straightforwardly generalized to the multivariate framework. The factor structure investigated in the previous section will be used to specify a new, parsimonious parameterization of volatility dynamics. It turns out that such a specification is most easily achieved in the stochastic volatility framework proposed by Taylor (1986), and Harvey, Ruiz and Shephard (1994).

Most of the multivariate ARCH models that have been developed introduce ad hoc restrictions on the number of parameters, and are mostly inadmissible in the application to exchange rates. An example is the constant conditional correlation assumption of Baillie and Bollerslev (1990). Consider the bivariate case with two exchange rates (*e.g.* dollar/yen and dollar/mark):

$$\Sigma_0(t) = \begin{pmatrix} \sigma_1^2(t) & \rho\sigma_1(t)\sigma_2(t) \\ \rho\sigma_1(t)\sigma_2(t) & \sigma_2^2(t) \end{pmatrix},$$

with $\sigma_1^2(t)$ and $\sigma_2^2(t)$ the variance of the dollar/yen and dollar/mark rate respectively. By changing the numeraire from the dollar to the yen we obtain the transformed "yen" covariance matrix

$$\Sigma_1(t) = \begin{pmatrix} \sigma_1^2(t) & \sigma_1^2(t) - \rho\sigma_1(t)\sigma_2(t) \\ \sigma_1^2(t) - \rho\sigma_1(t)\sigma_2(t) & \sigma_1^2(t) + \sigma_2^2(t) - 2\rho\sigma_1(t)\sigma_2(t) \end{pmatrix},$$

which no longer has the constant conditional correlation property. The same problem arises in the diagonal ARCH model of Bollerslev, Engle and Wooldridge (1988). The factor model of Diebold and Nerlove (1989) is also currency specific. Their model reads

$$s_{i0}(t) = \gamma_i v(t) + e_i(t),$$

where γ_i ($i = 1, \dots, n$) are factor loadings on the single factor $v(t)$, and all exchange rates are expressed with the dollar as numeraire. Diebold and Nerlove (1989) assume that all time varying volatility is due to the common factor, *i.e.* the dollar numeraire effect. The diagnostics in Table 4.3 indicated however, that all currency factors exhibit conditional

heteroskedasticity. We would therefore still need a multivariate ARCH model for $v(t)$ and all $e_i(t)$.

The FACTOR-ARCH model of Engle, Ng and Rothschild (1990) is also not directly applicable to a system of exchange rates. In our zero factor exchange rate model (4.4) we have $s_0(t) = \mathbf{P}e(t)$. The exchange rate model obviously has constant factor loadings, since the matrix $\mathbf{P} = (-\iota|\mathbf{I})$ is completely known. But since all currency specific factors can be (and probably are) heteroskedastic, the number of factors ($n+1$) is larger than the number of elements (n) in the vector time series $s_0(t)$. Specification of the GARCH structure for the factor variances is not trivial in that case, especially since the diagnostics in Table 4.3 also indicate cross effects from the volatility of one factor to the volatility of all other factors.

We opted to apply the multivariate stochastic volatility (MSV) model used by Harvey, Ruiz and Shephard (1994) (HRS hereafter), and respecify it to fit into the covariance structure of Section 4.1. The use of a stochastic volatility model is a natural choice, since the volatility is modelled as a latent variable. Each of the volatility components λ_i can be viewed upon as a latent factor too. Another advantage of MSV is that no restrictions on the parameters are needed to ensure that the volatility series are always positive. The theoretical and empirical issues of univariate stochastic volatility models were already set out in chapters 2 and 3. In the next subsection we describe our multivariate stochastic volatility model.

4.2.2 A Multivariate Stochastic Volatility Model for Exchange Rates

In our model $w(t)$ is a vector of length $\frac{1}{2}n(n+1)$ with typical element $w_k(t) = \ln\{y_k(t)\} = \ln\{s_{ij}^2(t)\}$, ($i=0, \dots, n-1; j=i+1, \dots, n$), *i.e.* we use all possible bilateral exchange rates, see Section 4.1. According to the factor model (4.4) the variance of any bilateral exchange rate s_{ij} is the sum of two currency specific variances. This implies that

$$\exp(h_{ij}(t)) = \lambda_i(t) + \lambda_j(t) = \exp(h_i(t)) + \exp(h_j(t)), \quad (4.7)$$

where $h_i(t)$ measures the specific volatility of currency i . Linearizing (4.7) around some \bar{h}_i and \bar{h}_j gives

$$\begin{aligned} h_{ij}(t) = & \ln\{\exp(\bar{h}_i) + \exp(\bar{h}_j)\} + \\ & \frac{\exp(\bar{h}_i)(h_i(t) - \bar{h}_i) + \exp(\bar{h}_j)(h_j(t) - \bar{h}_j)}{\exp(\bar{h}_i) + \exp(\bar{h}_j)} \end{aligned} \quad (4.8)$$

If we further assume a common point of linearization $\bar{h}_i = \bar{h}_j$ the linear approximation reduces to

$$h_{ij}(t) = \ln(2) + \frac{1}{2} (h_i(t) + h_j(t)). \quad (4.9)$$

Equation (4.9) shows that every bilateral exchange rate has two factors that define its volatility. In a system with $n+1$ currencies we thus have $n+1$ common volatility factors, which leads to the following structure for the measurement equations of the state space model

$$w(t) = \frac{1}{2} \mathbf{Z} h(t) + (\alpha + \ln(2)) \iota + \xi(t), \quad \text{Var}[\xi(t)] = \Phi,$$

where \mathbf{Z} has been defined in (4.4) and (4.5). To reduce the number of parameters in the estimation procedure we assumed that all exchange rates have the same type of distribution, implying that the diagonal elements of Φ are equal to ϕ^2 . Furthermore, we specified the correlation structure of $\xi(t)$ as

$$\text{Corr}(\xi_{ij}, \xi_{kl}) = \begin{cases} r & \text{if } i = (k \text{ or } l), \text{ or } j = (k \text{ or } l) \\ 0 & \text{otherwise.} \end{cases}$$

In this case Φ contains only two unknown parameters, and can be written as

$$\Phi = \phi^2 \begin{pmatrix} 1 & r & r & r & r & 0 \\ r & 1 & r & r & 0 & r \\ r & r & 1 & 0 & r & r \\ r & r & 0 & 1 & r & r \\ r & 0 & r & r & 1 & r \\ 0 & r & r & r & r & 1 \end{pmatrix}$$

If ϑ is the correlation between two elements of $\epsilon(t)$, then HRS derive the correlation between the corresponding elements of $\xi(t)$ as

$$r = \frac{2}{\pi^2} \sum_{n=1}^{\infty} \frac{(n-1)! \Gamma(\frac{1}{2})}{\Gamma(n + \frac{1}{2})}$$

The unconditional correlation between ϵ_{ij} and ϵ_{kl} is equal to $\frac{1}{2}$ under the assumption of equal factor variances. This implies that we expect r to be approximately equal to 0.11.

The state transition equation for a multivariate process is the first order vector autoregression

$$h(t) = (\mathbf{I} - \mathbf{A})\mu + \mathbf{A}h(t-1) + \eta(t),$$

where \mathbf{A} is a $((n+1) \times (n+1))$ matrix; μ is an $(n+1)$ vector of constants; and where $E[\eta(t)\eta(t)'] = \Omega$. The constant terms μ and α are not separately identified. We therefore reparameterize the system as

$$\begin{aligned} w(t) &= \frac{1}{2} \mathbf{Z} (x(t) + \zeta) + \xi(t) \\ x(t) &= \mathbf{A}x(t-1) + \eta(t) \end{aligned} \quad (4.10)$$

where $\zeta = \mu + \tilde{l}(\alpha + \ln(2))$, with \tilde{l} an $(n+1)$ vector of ones, and where $x(t) = h(t) - \mu$. Without restrictions on \mathbf{A} or ζ the total number of parameters is 32 in our application with the four major currencies.

Our model differs from the factor model of HRS, who define $w(t)$ as a vector of logarithms of squared exchange rate changes against a common numeraire currency. The typical element in this vector is $w_j(t) = \ln\{s_{0j}^2(t)\}$, with currency 0 being the common numeraire. The problem with this specification is that it is numeraire dependent. In squaring the exchange rates against a single numeraire one loses valuable information about the covariances. In order to construct a model for the whole system of exchange rates we need to augment the vector $w(t)$ to include the logarithm of the squares of all possible bilateral exchange rate changes, as described above and in accordance with the covariance structure investigated in section 4.1.

Using specification (4.10), the Kalman filter prediction equation provides a recursive formula for the conditional log variance $x(t)$ (up to a constant). Analogous to the univariate case the implied EGARCH type specification reads

$$\begin{aligned}\hat{x}(t+1|t) &= 2\mathbf{A}\mathbf{Z}' \left(\mathbf{Z}\bar{\mathbf{P}}\mathbf{Z}' + 4\Phi \right)^{-1} w(t) + \\ &\quad \mathbf{A} \left(\mathbf{I} - 2\bar{\mathbf{P}}\mathbf{Z}' \left[\mathbf{Z}\bar{\mathbf{P}}\mathbf{Z}' + 4\Phi \right]^{-1} \mathbf{Z} \right) \hat{x}(t|t-1) \\ &\equiv \mathbf{B}w(t) + \mathbf{C}\hat{x}(t|t-1)\end{aligned}$$

where $\bar{\mathbf{P}}$ is the solution to an algebraic Riccati equation. The conditional heteroskedasticity formulation also involves all squares and cross products of the log squared exchange rate changes. However, the number of free parameters in \mathbf{B} and \mathbf{C} is restricted.

4.3 Stochastic Volatility Results

The data for the empirical results in this section is

$$w_{ij}(t) = \ln \left\{ \left(\Delta \ln\{S_{ij}(t)\} - \overline{S_{ij}(t)} \right)^2 \right\},$$

where $\overline{S_{ij}(t)}$ is the sample mean of $\Delta \ln\{S_{ij}(t)\}$. A set of descriptive statistics for the six transformed bilateral exchange rates is reported in Table 4.4. The transformed series show considerable autocorrelation; especially the autocorrelations of the dollar/yen and dollar/pound exchange rates are very persistent and still sizable after 30 lags. The differences between these two autocorrelation patterns and the other four give an indication that one time-varying factor is not enough to describe the dynamics of the currencies in our sample.

TABLE 4.4: Summary Statistics of Transformed Exchange Rate Changes $w_{ij}(t)$

| w_{ij} | JP/US | GE/US | UK/US | GE/JP | UK/JP | UK/GE |
|---|------------------|------------------|-----------------|-----------------|-----------------|-----------------|
| Mean | -1.25 | -0.79 | -1.08 | -1.01 | -1.06 | -1.43 |
| Std. Dev. | 2.73 | 2.42 | 2.58 | 2.32 | 2.52 | 2.37 |
| Skewness | -1.14 | -1.24 | -1.03 | -1.21 | -1.09 | -0.86 |
| Kurtosis | 2.04 | 1.60 | 1.20 | 2.31 | 1.48 | 0.93 |
| Normality | 376.77* | 310.70* | 226.41* | 448.79* | 278.39* | 152.51* |
| Minimum | -15.11 | -12.00 | -14.33 | -13.25 | -10.49 | -11.44 |
| Maximum | 4.87 | 4.31 | 4.04 | 3.61 | 4.19 | 3.61 |
| Ljung-Box(30) | 510.37* | 127.81* | 459.97* | 86.00* | 93.57* | 59.91* |
| Autocorrelations | | | | | | |
| 1 | 0.20* | 0.09* | 0.17* | 0.08* | 0.07* | 0.12* |
| 2 | 0.15* | 0.12* | 0.19* | 0.06 | 0.10* | 0.10* |
| 3 | 0.16* | 0.10* | 0.16* | 0.11* | 0.07* | 0.09* |
| 4 | 0.15* | 0.07* | 0.20* | 0.03 | 0.08* | 0.07* |
| 5 | 0.14* | 0.10* | 0.14* | 0.07* | 0.03 | 0.02 |
| 10 | 0.11* | 0.04 | 0.15* | -0.01 | 0.04 | 0.07* |
| 15 | 0.12* | 0.03 | 0.14* | -0.02 | 0.05 | 0.03 |
| 20 | 0.10* | 0.05 | 0.08* | 0.09* | 0.04 | -0.01 |
| 25 | 0.11* | 0.02 | 0.08* | 0.03 | 0.07* | -0.00 |
| 30 | 0.09* | 0.03 | 0.06 | 0.00 | -0.04 | 0.02 |
| GARCH(1,1): $s_{ij}(t) = \delta + \nu_{ij}(t)$; $E_{t-1}[\nu_{ij}^2(t)] \equiv h_t = \alpha_0 + \alpha_1 h(t-1) + \alpha_2 \nu_{ij}^2(t-1)$ | | | | | | |
| δ | -0.075 (0.04) | -0.082 (0.04) | 0.037 (0.04) | 0.027 (0.04) | 0.131 (0.05) | 0.075 (0.04) |
| α_0 | 0.009 (0.01) | 0.020 (0.02) | 0.085 (0.04) | 0.118 (0.18) | 0.103 (0.10) | 0.172 (0.44) |
| α_1 | 0.044 (0.01) | 0.069 (0.02) | 0.079 (0.02) | 0.179 (0.14) | 0.059 (0.04) | 0.101 (0.13) |
| α_2 | 0.952 (0.01) | 0.924 (0.03) | 0.881 (0.02) | 0.763 (0.22) | 0.889 (0.07) | 0.774 (0.44) |

Notes: Variables are defined as $w_{ij}(t) = \ln\{[\Delta \ln\{S_{ij}(t)\} - \bar{S}_{ij}(t)]^2\}$. See Table 4.3 for definitions of statistics. Standard errors are between parentheses for GARCH(1,1) parameters. A '*' denotes significance at the 5% level. For autocorrelation tests the critical value is $\pm 2/\sqrt{N}$ with $N=965$.

Furthermore, for all exchange rates the $w_{ij}(t)$ are negatively skewed. This is an indication that the exchange rates are not as heavily fat-tailed as implied by the Cauchy distribution. The negative outliers in the data are due to many small changes in the exchange rates. The lower part of Table 4.4 shows parameter estimates for six univariate GARCH(1,1) models. Two exchange rates (dollar/yen and dollar/mark) turn out to be almost IGARCH, while the other four are much less persistent. The clear differences in the dynamics provides additional evidence that there are more factors than just the dollar.

The system (4.10) has been estimated by numerical optimization of the quasi-likelihood function obtained by assuming normality for $\xi(t)$ and $\eta(t)$. Parameter estimates are in Table 4.5; diagnostics in Table 4.6. The estimates of ϕ and the residual characteristics provide information about the type of distribution for $\epsilon(t)$ and $\xi(t) = \ln\{\epsilon(t)^2\}$. The error variance in the measurement equations is estimated very precisely and almost equal to π^2 , which is very close to the value implied by a Cauchy distribution for exchange rate innovations, and much larger than what is implied by the log chi-squared distribution. The prediction errors are less negatively skew than is implied by the log chi-squared. The results seem to indicate that a Student- t distribution with low degrees of freedom might be a good choice for $\epsilon(t)$. The dynamic specification do not completely describe the mark volatility, as the diagnostics for the prediction errors involving the mark imply significant residual autocorrelation.

The correlation parameter r was estimated freely but corresponds to the theoretical value ($r \approx 0.11$) derived from the formulas in HRS, supporting the approximation $\bar{h}_i = \bar{h}_j$ that we used in the linearization of (4.8).

The structure of the error covariance matrix Ω of the transition equations implies that the innovations of all four variance components are highly positively correlated. Given our use of weekly data, it means that an increase in the volatility of one currency gets transmitted to increased volatility in all other currencies within a week. It also means that, for example, an increase in the volatility of dollar exchange rates leads to increased volatility in the mark/pound cross rate. Although the weekly innovations are highly correlated, in the long run the four variance components behave very differently (see Table 4.5). High volatility of the mark tended to go together with low volatility of the dollar and yen. This is in close agreement with the stylized facts for different subperiods in Table 4.1.

The differences between the innovations covariance structure Ω and the unconditional covariance matrix are caused by the large negative estimates for some elements in the transition matrix \mathbf{A} . Despite the negative off-diagonal elements the system has three large eigenvalues (see Table 4.5) that are close to unity, indicating that the volatility series $h(t)$

TABLE 4.5: Parameter Estimates of Stochastic Volatility Model

| | | | | |
|--|--|------------------|--------------------------------|------------------|
| | $w(t) = \frac{1}{2}\mathbf{Z}(x(t) + \zeta) + \xi(t),$ | | $\text{Var}[\xi(t)] = \Phi$ | |
| | $x(t) = \mathbf{A}x(t-1) + \eta(t),$ | | $\text{Var}[\eta(t)] = \Omega$ | |
| | dollar | yen | mark | pound |
| ζ_i | -1.018 (0.47) | -1.152 (0.29) | -0.100 (0.18) | -1.384 (0.25) |
| $a_{\text{dollar},j}$ | 0.820 (0.10) | -0.266 (0.14) | -0.871 (0.31) | 0.109 (0.14) |
| $a_{\text{yen},j}$ | 0.122 (0.11) | 0.541 (0.15) | -1.019 (0.33) | 0.146 (0.17) |
| $a_{\text{mark},j}$ | -0.080 (0.05) | -0.146 (0.06) | 0.496 (0.14) | 0.032 (0.07) |
| $a_{\text{pound},j}$ | -0.062 (0.06) | -0.143 (0.10) | -0.449 (0.25) | 0.980 (0.08) |
| Volatility innovation covariance matrix $\Omega = \text{Var}[\eta(t)]$ | | | | |
| dollar | 1.183 | 0.957 | 0.510 | 0.570 |
| yen | <i>0.783</i> | 1.263 | 0.446 | 0.569 |
| mark | <i>0.842</i> | <i>0.713</i> | 0.310 | 0.361 |
| pound | <i>0.736</i> | <i>0.711</i> | <i>0.913</i> | 0.507 |
| Unconditional volatility $\bar{\Omega} = \mathbf{A}\bar{\Omega}\mathbf{A}' + \Omega$ | | | | |
| dollar | 4.781 | 2.374 | -0.458 | 1.262 |
| yen | <i>0.627</i> | 2.999 | -0.339 | 1.010 |
| mark | <i>-0.245</i> | <i>-0.229</i> | 0.728 | 0.239 |
| pound | <i>0.437</i> | <i>0.442</i> | <i>0.212</i> | 1.742 |
| Standard deviation of $\xi(t)$: $\phi = 3.148$ (0.037) | | | | |
| Correlation $\xi_i(t), \xi_j(t)$: $r = 0.130$ (0.012) | | | | |
| Roots of system (eigenvalues of \mathbf{A}): | | | | |
| | 0.979 | 0.929 | 0.875 | 0.054 |

Notes: Exchange rates are expressed in percentages per week. Standard errors are in parentheses. *Italics* in the lower triangle of a covariance matrix denote correlations. All eigenvalues of \mathbf{A} are real.

TABLE 4.6: Diagnostics of Stochastic Volatility Model

$$\xi(t) = w(t) - \frac{1}{2}\mathbf{Z}(\hat{x}(t|t-1) + \zeta)$$

| ξ_{ij} | JP/US | GE/US | UK/US | GE/JP | UK/JP | UK/GE |
|---------------|--------|--------|--------|--------|--------|--------|
| Skewness | -1.43 | -0.94 | -0.94 | -1.44 | -1.08 | -1.44 |
| Kurtosis | 3.17 | 1.72 | 1.91 | 3.63 | 1.86 | 1.17 |
| Normality | 729.9* | 261.4* | 289.0* | 860.2* | 326.4* | 448.7* |
| Ljung-Box(10) | 8.67 | 20.55* | 17.82 | 19.62* | 5.40 | 27.77* |
| Ljung-Box(20) | 16.38 | 37.25* | 30.26 | 28.69 | 15.54 | 48.29* |
| ARCH(1) | 0.34 | 5.57* | 0.77 | 0.17 | 0.40 | 1.02 |

Notes: This table provides diagnostics of the Kalman filter prediction errors. See Table 4.3 for explanatory notes.

might be integrated. Further analysis of the dynamic implications is reported in a set of causality tests in Table 4.7. There seems to be no lagged relations from the dollar volatility to volatility in any of the other currencies: dollar news is transmitted within a week. The yen and mark, however, strongly influence all the other currencies. Especially the large negative elements of the mark column in **A** are noticeable (see Table 4.5).

The causality pattern is consistent with the results of Engle, Ito and Lin (1990), who used a dataset with four observations a day for the dollar/yen exchange rate. They can identify the separate country specific news because they have observations on the opening and closing prices of different markets. Strong volatility spillovers were found from Japan to the U.S., and vice versa. Their results seem to suggest that *'the Tokyo news has a greater impact on the volatility spillovers'* and *'the volatility in the Tokyo market (...) had a great impact on the world volatility'* (Engle, Ito and Lin (1990, pp. 535, 538)). Both these facts are related to our finding of a strong lagged effect from yen volatility to volatility in the other currencies. Interestingly, we find the same effect for the mark, but not for the pound and dollar.

Figures 4.1 and 4.2 contain time series plots of the series for each of the elements of the vectors $x(t|t-1)$ and $x(t|T)$, respectively. These series are obtained from the standard Kalman prediction and smoother recursions. The conditional log volatility $x(t|t-1)$ is used in the EGARCH representation of the stochastic volatility model. It is much smoother than the series $x(t|T)$ that attempts to add an estimate of $x(t)$ based on all information in the sample. The dollar volatility series shows more fluctuations than the other ones, especially

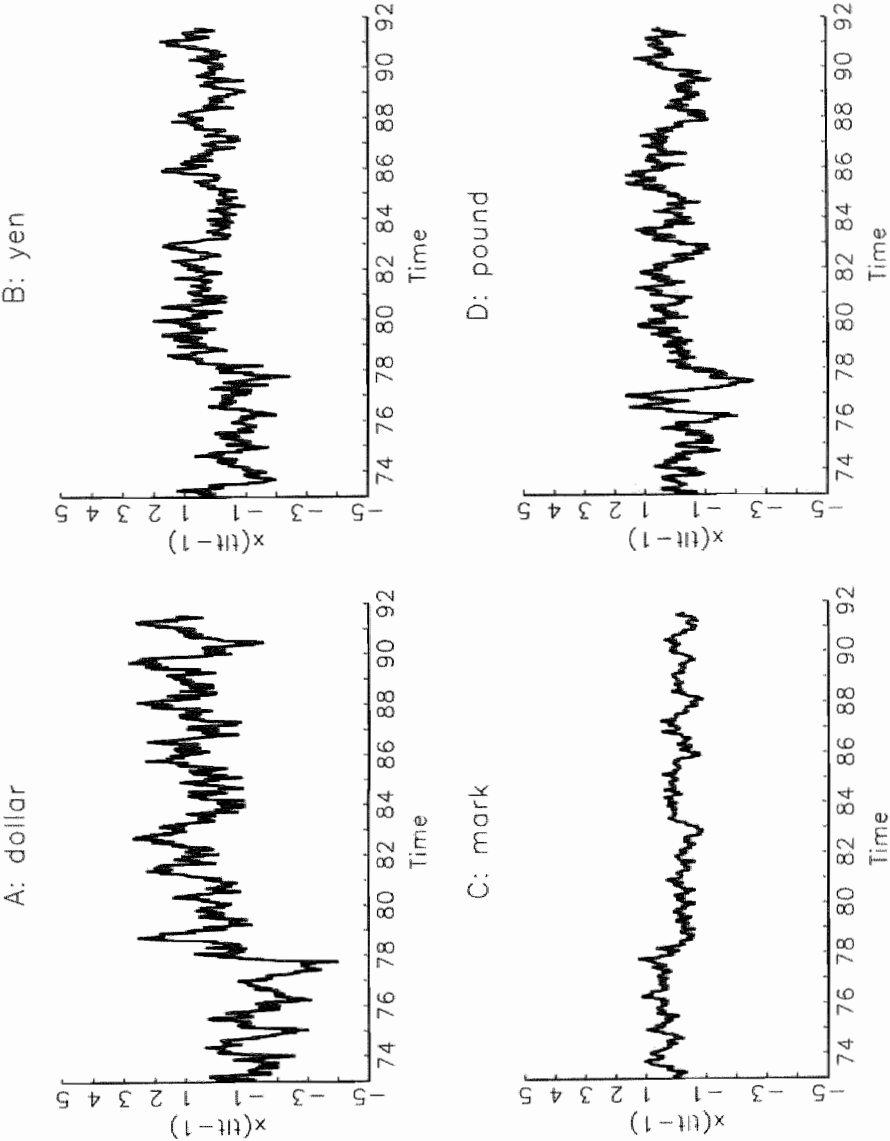


FIGURE 4.1: Time Series of Estimated Conditional Log-Variances

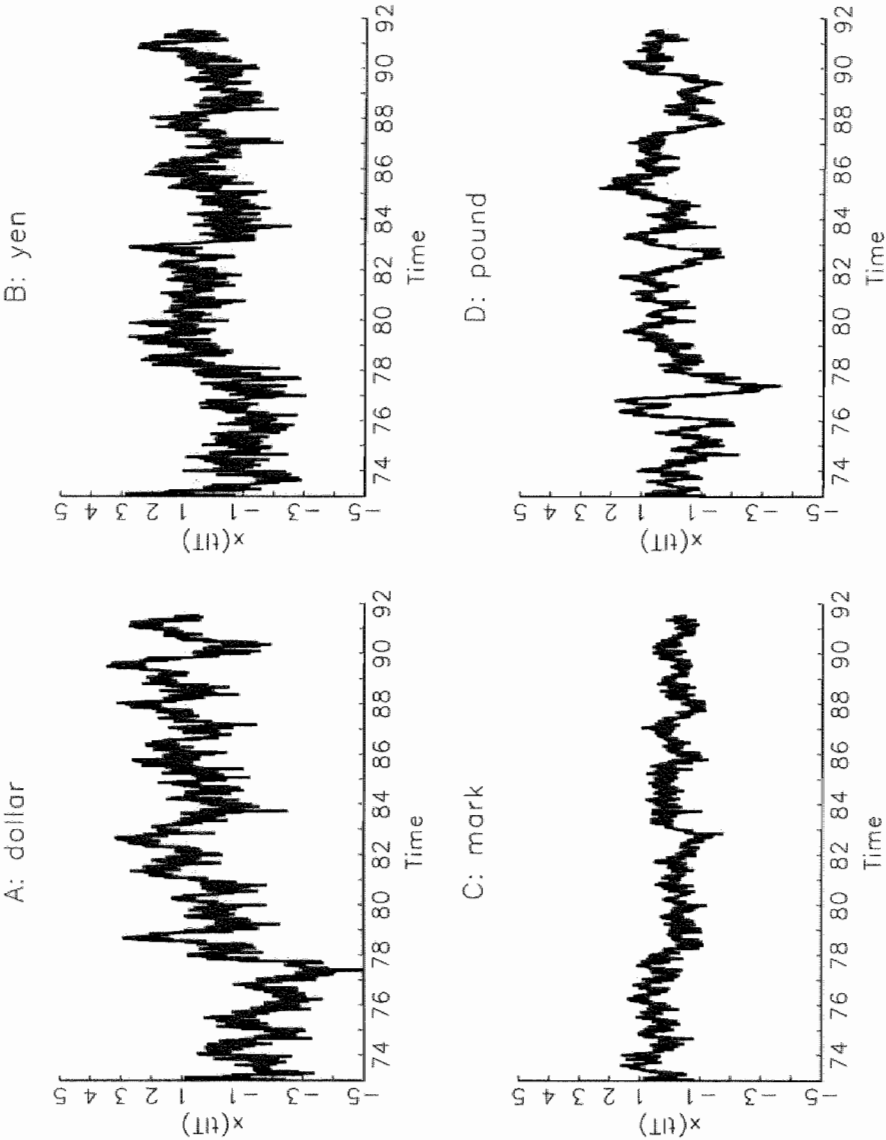


FIGURE 4.2: Time Series of "Smoothed" Estimates of the Log-Volatilities

TABLE 4.7: Stochastic Volatility Tests

| $w(t) = \frac{1}{2}\mathbf{Z}(x(t) + \zeta) + \xi(t), \quad \text{Var}[\xi(t)] = \Phi$ $x(t) = \mathbf{A}x(t-1) + \eta(t), \quad \text{Var}[\eta(t)] = \Omega$ | | | |
|---|--|--------|----|
| Hypothesis | | Wald | df |
| (1) | Equal unconditional variances: $\zeta_1 = \zeta_2 = \zeta_3 = \zeta_4$ | 2.78 | 3 |
| (2) | Diagonal dynamics: $a_{ij} = 0 \ (i \neq j)$ | 69.02* | 12 |
| (3) | Causality (columnwise tests): | | |
| | dollar to other currencies: $a_{21} = a_{31} = a_{41} = 0$ | 1.59 | 3 |
| | yen to other currencies : $a_{12} = a_{32} = a_{42} = 0$ | 13.38* | 3 |
| | mark to other currencies : $a_{13} = a_{23} = a_{43} = 0$ | 14.17* | 3 |
| | pound to other currencies : $a_{14} = a_{24} = a_{34} = 0$ | 2.17 | 3 |
| (4) | Causality (rowwise tests): | | |
| | other currencies to dollar: $a_{12} = a_{13} = a_{14} = 0$ | 9.10* | 3 |
| | other currencies to yen : $a_{21} = a_{23} = a_{24} = 0$ | 15.71* | 3 |
| | other currencies to mark : $a_{31} = a_{32} = a_{34} = 0$ | 11.77* | 3 |
| | other currencies to pound : $a_{41} = a_{42} = a_{43} = 0$ | 9.91* | 3 |
| (5) | Diagonal variance innovations: $\omega_{ij} = 0 \ (i \neq j)$ | 146.5* | 6 |

Notes: *Wald* is the Wald test statistic for the hypothesis in the first column. The covariance matrix of the parameters is computed from the outer product of the scores of the quasi log-likelihood function. An asterisk (*) denotes significance at the 5% level using the $\chi^2(df)$ table.

when compared to the relatively constant Deutsche mark (log-) volatility. The sub-period averages of $x(t)$ are consistent with the estimates in Table 4.1: the dollar volatility is moderate until the end of 1977 compared to the eighties; yen volatility is high in the late seventies; and mark volatility is slowly but steadily decreasing over the sample. It is hard to make any reliable inference on the volatility at a particular point in time, since the standard errors of the state vector elements $x_i(t|t-1)$ and $x_i(t|T)$ obtained from the steady state Kalman recursions are large, even conditional on the parameter estimates.

4.4 Implications for Risk Premia

From a practical viewpoint the factor model of exchange rate volatilities provides useful information on the riskiness of investing funds internationally. Each of the currency factors

can be interpreted as a measure for the volatility of country-specific news that affects that currency alone. Consequently, an investor can monitor the currency risk of investing in that country more accurately. An important application could be the hedging of currency exposure: a higher volatility on the news factor would lead to higher hedge ratios *ceteris paribus*. Alternatively, the currency factors could be used to investigate the existence and magnitude of risk premia. In this section we will turn to this application.

It is a well-known fact in the exchange rate literature that the forward rate is a biased estimator for the future spot rate. Maintaining rational expectations, the combined existence of an efficient foreign exchange market and a time-varying risk premium can account for the bias. The literature on risk premia has concentrated on developing models to disentangle the joint hypothesis of market efficiency and constant risk premia. A few references in this field are Hodrick (1987), Bollerslev, Chou and Kroner (1992), and Bekaert, Hodrick and Marshall (1993). Bekaert and Hodrick (1993) provide an extensive econometric investigation on the existence of the risk premium for the major currencies in the presence of conditional heteroskedasticity. Survey data provide direct evidence on risk premia, see *e.g.* Frankel and Froot (1987).

The simplest model in this field is the model of Domowitz and Hakkio (1985) (DH from now on). DH base their model on the asset pricing theory of Lucas (1982). In the DH model the risk premium is related to the difference between the conditional variances of the money supply in two countries. In the empirical part of their paper DH investigate if there is any evidence of a time-varying risk premium in several dollar exchange rates. DH analyze univariate cases and, due to lack of identification, assume that only the dollar news term is heteroskedastic. In a multivariate setup this simplifying assumption is unnecessary. The covariance between exchange rate changes helps to identify all three news components, and thus their variances, in a bivariate system of exchange rates. The identification is brought upon by the existence of a common U.S. news factor in a system of dollar exchange rates. Our extension to the multivariate case differs from Baillie and Bollerslev (1990) due to the explicit use of the currency specific factors.

Extending the risk premium model of DH to a model with the factor structure of the previous sections, leads to the following expression for the one period risk premium:

$$RP_{ij}(t) = \frac{1}{2} (\lambda_i(t) - \lambda_j(t)),$$

where $\lambda_i(t) \equiv \exp(h_i(t|t-1))$ is the conditional variance of factor i . Our dataset consists of weekly exchange rate changes, whereas the most commonly studied risk premium is the one implicit in forward contracts with a maturity of one month. The one month risk

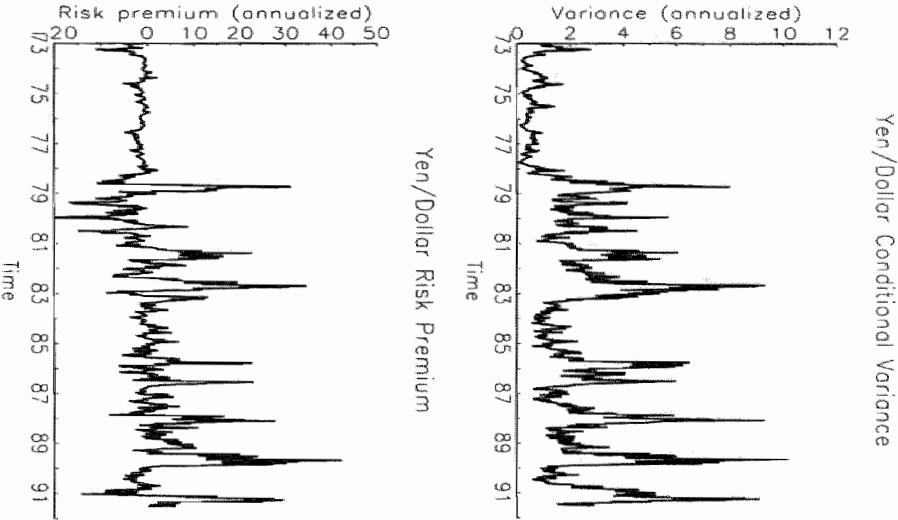


FIGURE 4.3: Conditional Variance of yen/dollar Exchange Rate $\lambda_{JP}(t,4)+\lambda_{US}(t,4)$ and the risk premium $\frac{1}{2}(\lambda_{US}(t,4) - \lambda_{JP}(t,4))$.

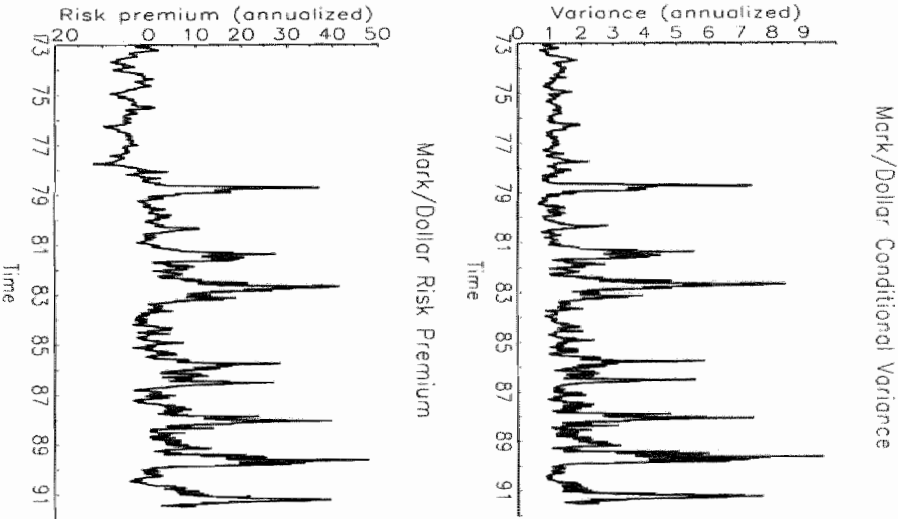


FIGURE 4.4: Conditional Variance of mark/dollar Exchange Rate $\lambda_{GE}(t,4)+\lambda_{US}(t,4)$ and the risk premium $\frac{1}{2}(\lambda_{US}(t,4) - \lambda_{GE}(t,4))$.

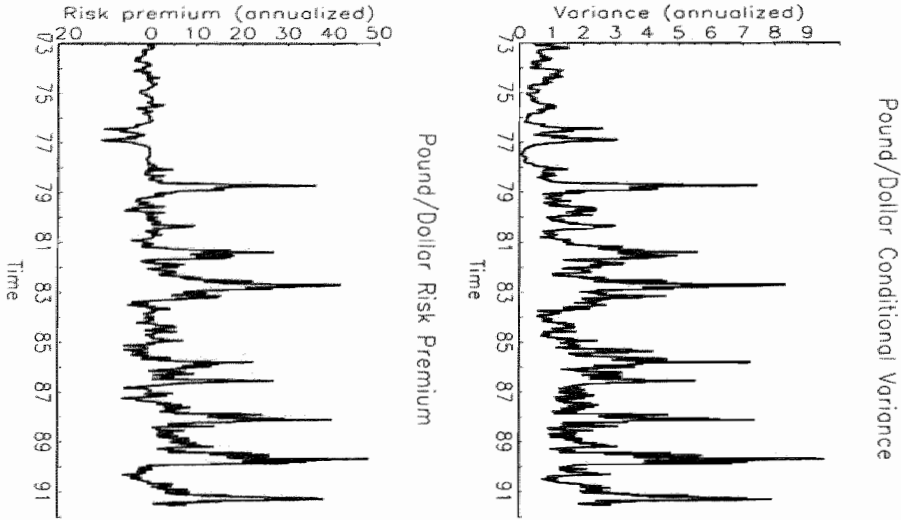


FIGURE 4.5: Conditional Variance of pound/dollar Exchange Rate $\lambda_{UK}(t, 4) + \lambda_{US}(t, 4)$ and the risk premium $\frac{1}{2}(\lambda_{US}(t, 4) - \lambda_{UK}(t, 4))$.

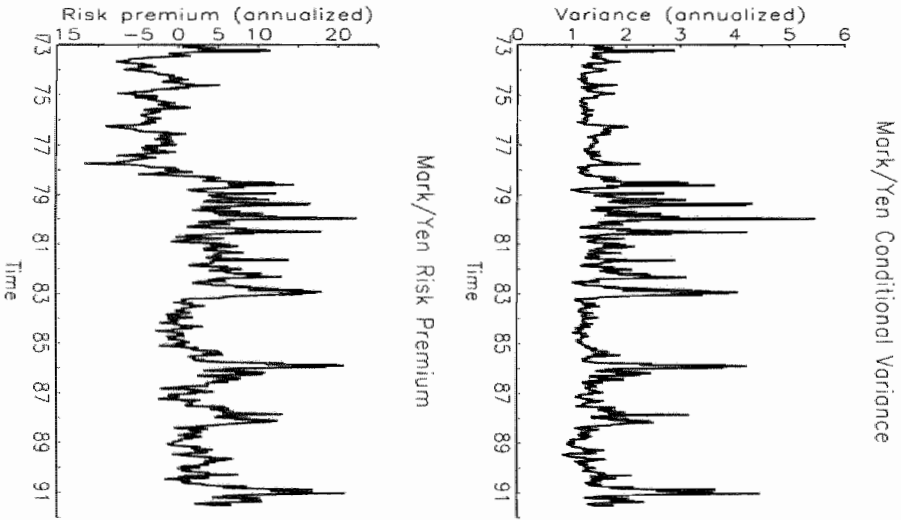


FIGURE 4.6: Conditional Variance of mark/yen Exchange Rate $\lambda_{GE}(t, 4) + \lambda_{JP}(t, 4)$ and the risk premium $\frac{1}{2}(\lambda_{JP}(t, 4) - \lambda_{GE}(t, 4))$.

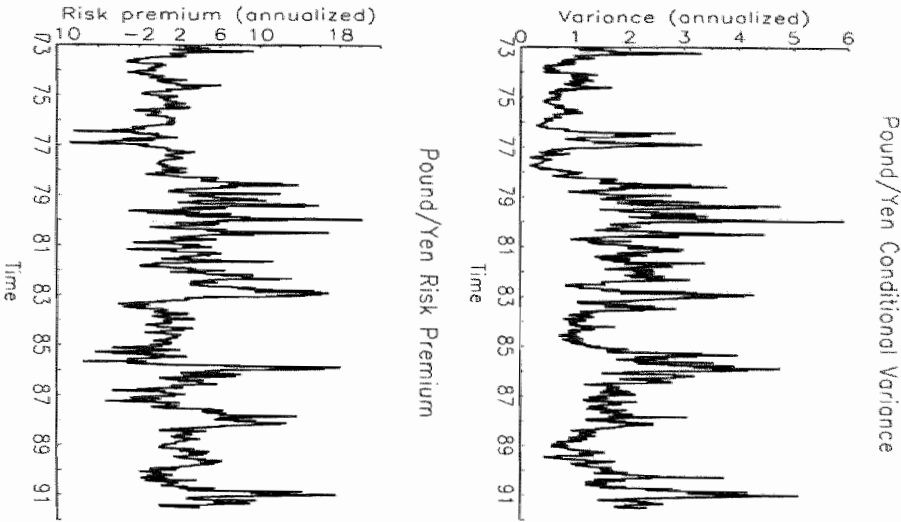


FIGURE 4.7: Conditional Variance of pound/yen Exchange Rate $\lambda_{UK}(t,4)+\lambda_{JP}(t,4)$ and the risk premium $\frac{1}{2}(\lambda_{JP}(t,4) - \lambda_{UK}(t,4))$.

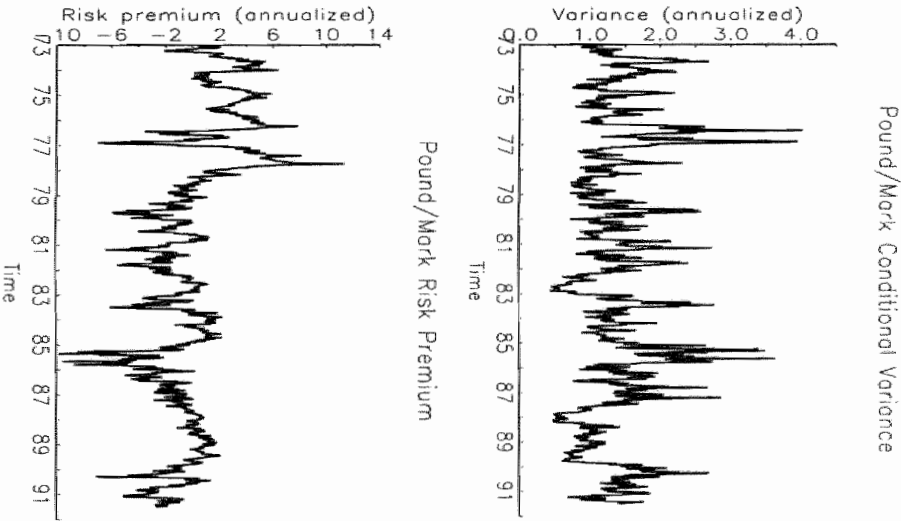


FIGURE 4.8: Conditional Variance of pound/mark Exchange Rate $\lambda_{UK}(t,4)+\lambda_{GE}(t,4)$ and the risk premium $\frac{1}{2}(\lambda_{GE}(t,4) - \lambda_{UK}(t,4))$.

premium is computed using the expected variance over a period of four weeks:

$$\lambda_i(t, 4) = C_{i4} \sum_{k=1}^4 \exp \left(\mathbf{g}_i' \mathbf{A}^k \hat{x}(t + k|t) \right),$$

where \mathbf{g}_i is an indicator vector with zeros in every row except in row i , which is one; and where C_{i4} is a constant arising from taking expectations of a log-normal random variable, and also related to the constant terms α and μ . Because α and μ are not separately identified we have estimated the scale factor implicitly by requiring that the sample average of $\lambda_i(t)$ is equal to the full sample estimation of the constant λ_i in Table 4.1. The resulting expression for the risk premium of the exchange rate between countries i and j is

$$\widehat{RP}_{ij}(t) = \frac{1}{2} (\lambda_i(t, 4) - \lambda_j(t, 4)).$$

Figures 4.3 through 4.8 show the estimated risk premia and also the conditional variances of the six bilateral exchange rates, which were the original proxy for the risk premium used by DH. Table 4.8 provides some summary statistics of the time series of the risk premia.

For the three dollar exchange rates the risk premium is dominated by the U.S. news component after 1980. In the seventies the risk premia on the dollar were much smaller and fluctuated less. The sign of the dollar risk premium also differs between the seventies and eighties. For example, the pound/dollar rate (Figure 4.5) shows two troughs in 1976 that correspond to peaks in the variance of the pound/dollar rate. The same phenomenon shows up in the yen/dollar rate in 1973. In the eighties the signs are reversed: all peaks in the conditional variance of the dollar rates are also peaks in the risk premia.

The figures and Table 4.8 indicate that the estimated risk premia fluctuate substantially. On average the size of the risk premium is small relative to its own standard deviation for all the subperiods, and almost negligible compared to the variance of exchange rate innovations (see tables 4.1 and 4.2 and note the difference in units). The conditional variances in figures 4.3 – 4.8 appear much more irregular than the log-volatility in Figure 4.1. Taking the exponent of the conditional log variances attenuates the upward movements in the conditional variance plots during periods of high (logarithmic) volatility.

Our measures of the risk premium can be compared with results from studies that use survey data. With survey data the risk premium is directly observable as the difference between the expected future spot rate and the forward rate for the same horizon. Frankel and Froot (1987), for instance, found that the risk premium varied between 2 and 10 percent on an annual basis. Similar magnitudes are reported in Cavaglia, Verschoor and Wolff (1994). From Table 4.8 it can be seen that the average risk premia are of the same magnitude. From the plots and the table we see that risk premia show a large amount of

TABLE 4.8: Summary Statistics of Risk Premia

| | JP/US | GE/US | UK/US | GE/JP | UK/JP | UK/GE |
|---|--------|--------|--------|--------|-------|-------|
| January 1973 - June 1991 (Full sample) | | | | | | |
| Mean | 1.04 | 3.59 | 3.53 | 2.55 | 2.50 | -0.05 |
| Std. dev. | 7.05 | 8.80 | 8.04 | 5.17 | 4.05 | 2.80 |
| Minimum | -19.88 | -12.00 | -10.86 | -11.57 | -8.79 | -9.81 |
| Maximum | 42.18 | 47.96 | 47.85 | 22.45 | 20.16 | 11.36 |
| January 1973 - December 1976 | | | | | | |
| Mean | -1.60 | -3.89 | -1.56 | -2.29 | 0.04 | 2.33 |
| Std. dev. | 1.57 | 2.36 | 2.27 | 3.13 | 2.56 | 2.43 |
| Minimum | -11.07 | -9.68 | -10.86 | -9.03 | -8.79 | -7.03 |
| Maximum | 2.07 | 2.16 | 2.80 | 11.39 | 9.29 | 7.86 |
| January 1977 - December 1980 | | | | | | |
| Mean | -2.34 | 1.12 | 1.67 | 3.45 | 4.00 | 0.55 |
| Std. dev. | 6.60 | 6.74 | 5.62 | 5.89 | 4.01 | 3.12 |
| Minimum | -19.88 | -12.00 | -5.78 | -11.57 | -3.04 | 5.87 |
| Maximum | 31.09 | 37.06 | 36.20 | 22.45 | 20.16 | 11.36 |
| January 1981 - September 1985 | | | | | | |
| Mean | 2.53 | 6.24 | 4.84 | 3.71 | 2.31 | -1.40 |
| Std. dev. | 6.72 | 8.07 | 8.32 | 4.16 | 4.37 | 2.30 |
| Minimum | -9.05 | -3.47 | -5.85 | -2.67 | -7.52 | -9.81 |
| Maximum | 34.49 | 41.61 | 41.72 | 18.05 | 16.84 | 2.22 |
| September 1985 - February 1987 (Plaza - Louvre) | | | | | | |
| Mean | 2.31 | 7.98 | 5.54 | 5.66 | 3.23 | -2.43 |
| Std. dev. | 5.85 | 7.17 | 6.63 | 5.12 | 4.62 | 1.64 |
| Minimum | -6.17 | -3.21 | -5.78 | -2.09 | -4.66 | -8.77 |
| Maximum | 22.87 | 28.56 | 27.08 | 20.81 | 17.99 | 0.19 |
| February 1987 - June 1991 | | | | | | |
| Mean | 4.51 | 8.66 | 8.02 | 4.14 | 3.51 | -0.63 |
| Std. dev. | 8.93 | 9.93 | 10.13 | 4.04 | 3.61 | 1.63 |
| Minimum | -14.35 | -4.14 | -6.16 | -2.43 | -5.28 | -7.07 |
| Maximum | 42.18 | 47.96 | 47.85 | 20.92 | 17.58 | 2.08 |

Notes: Units are annualized percentages of risk premium on one month forward contracts.
 A positive entry signifies that the numeraire in the column heading is the more risky currency.

variability over the whole sample. We find that weekly risk premia sometimes amount to 40 percent per year. Fama (1984) shows that the variability of the risk premium should be larger than the conditional variance of exchange rate changes. The empirical finding that the regression coefficient of a regression of exchange rate changes on the forward premium is usually negative together with rational expectations implies that the covariance between the expected rate of depreciation and the risk premium must be negative and larger in absolute value than the variance of the expected rate of depreciation. Thus, the variability of the risk premium is large enough to make the forward premium predict the wrong direction for the expected rate of exchange rate changes.²

4.5 Conclusions

All bilateral exchange rates, expressed vis-a-vis a common numeraire currency, contain at least one common factor due to the numeraire effect. We have examined empirically to what extent the movements among the four major currencies can be explained by just a set of currency specific factors, each representing the specific news in one of the currencies (dollar, yen, mark, pound). For the seventies we find that all currencies were approximately equally volatile, with some short periods of high German or high Japanese volatility. During the eighties the volatility of the U.S. dollar was dominant.

This factor structure has been used to specify a parsimonious multivariate model of time varying volatility. The resulting model is an extension of the stochastic variance model of Harvey, Ruiz and Shephard (1994). The conditional variances from the stochastic variance factor model follow approximately a highly restricted multivariate EGARCH process.

Using weekly data for the full floating exchange rate period 1973-1991, it appears that changes in dollar volatility quickly spread to changes in the volatility of other currencies, even affecting the volatility of cross exchange rates. The effects of increased yen and mark volatility take much longer to transmit to increases in the volatility of other currencies. The German mark appears to be the most stable major currency of the last two decades. Its average volatility is below that of the other major free floating currencies, and its volatility has been relatively constant over time.

The variance decomposition of exchange rates also provides a new approach to estimate foreign exchange risk premia in a complete system of currencies. We find that risk premia fluctuate considerably over the sample period.

²Economic explanations for the large negative correlation have been given by Hodrick and Srivastava (1986) using the Lucas (1982) model.

Chapter 5

Volatility and Volume: Balancing the Scales

The relationship between the conditional volatility of stock returns and the associated trading volume has been on the research agenda of financial economics for more than two decades now. Clark (1973) started the theoretical analysis of stock price movements and trading volume by introducing the so called Mixture of Distributions Hypothesis, which explores the idea of using trading volume as a proxy for a stochastic process of information arrival. This idea has been extended and refined in Copeland (1976) who uses sequential information processing, Epps and Epps (1976), who use another mixture, and Tauchen and Pitts (1983), who point out the possible problems in the simultaneous determination of volume and volatility. More recently Ghysels and Jasiak (1995) introduced the idea of time-deformation into these analyses. The empirical pendants of these theoretical analyses all find a strong contemporaneous correlation between volume and volatility movements, see Harris (1987), Smirlock and Starks (1988) and Lamoureux and Lastrapes (1990), among others. According to Laux and Ng (1993), however, the real test of the Mixture of Distributions Hypothesis, is whether lagged trading volume is related to current volatility. Another branch in this strand of the literature is centered around the market-microstructure models of Kyle (1985), Admati and Pfleiderer (1988), and Campbell, Grossman and Wang (1993) where the focus is on short-lived intraday or over-night relationships between returns and volumes.

In the econometrics literature on volatility and volume, the main topics of analysis are the possible non-linearities and asymmetries in financial return series. Volume-volatility relationships have been analyzed by Gallant, Rossi and Tauchen (1992, 1993), Tauchen, Zhang and Liu (1993), Andersen (1994a), Lamoureux and Lastrapes (1994) and LeBaron (1993). The parametric techniques used in the econometric volume-volatility literature

vary from adding exogenous volume to an ARCH model (Lamoureux and Lastrapes, 1990) to combining the Mixture of Distributions Theory with a SV model (Andersen, 1994a). Furthermore various semi-nonparametric (Gallant, Rossi and Tauchen, 1992, 1993), and completely non-parametric approaches (Tauchen, Zhang and Liu, 1993) have been used to investigate the interrelationships between stock price volatility and trading volume. The general upshot from the literature is that (a) the Mixture of Distributions Hypothesis at least requires some modification (Andersen, 1994a; Lamoureux and Lastrapes, 1994), (b) the leverage effect at best is present but short-lived at the individual asset level (Gallant, Rossi and Tauchen, 1993; Engle and Lee, 1993; Tauchen, Zhang and Liu, 1993), and (c) trading volume responds in a non-linear fashion to price shocks.² One issue that, in our view, needs more attention is the role of volume. Can it be used as a proxy for stochastic information arrival, is it the source of conditional heteroskedasticity, or has it yet another role? This chapter tries to shed some light on these questions.

Bauer and Nieuwland (1994) investigated a multiplicative relationship between volume and the conditional variance of stock returns, thereby using volume as a scale factor for volatility (in the sense of Brenner, Harjes and Kroner (1994), Koedijk, Nissen, Schotman and Wolff, 1994). Bauer and Nieuwland (1994) found significant parameter estimates for volume, which stresses the role volume plays in stock market dynamics. Laux and Ng (1993) use a similar approach to investigate the sources of GARCH in the futures market. They hypothesize that the time-varying rate of information arrival leads to the volatility dependencies captured by GARCH models and discriminate between systematic and unique volatility. In their multivariate approach they show that autocorrelation in the rate of information arrival can explain much of the persistence in unique volatility in the futures market, systematic effects however remain unexplained by the behaviour of the rate of information arrival. Recently Lamoureux and Lastrapes (1994) extended their approach by presenting a mixture model where trading volume is endogenized. Simulation results indicate that merely accounting for serial dependence in the process of information arrival does not eliminate volatility persistence of the GARCH form.

In this chapter we propose a new role for volume. A Stochastic Volatility (SV) model is developed in which volume acts as a scale factor for the conditional volatility. This model is motivated by the Mixture of Distributions Hypothesis. We apply Quasi Maximum Likelihood to estimate these models. Considerable attention is given to asymmetric effects from shocks in stock returns to (conditional) volatility. This so-called leverage effect appears to be difficult to extract from the series in our sample, especially when volume is taken into account.

²The leverage effect predicts that positive and negative shocks in returns have a different impact on the volatility. See Section 5.2.

The remainder of this chapter is organized as follows. In Section 5.1 we will discuss the mixture of distributions hypothesis in more detail. We will show that it can be linked with a SV model. In this section we will also present the basic model that we use in the empirical section. The data are described in Section 5.2. We use data on five individual stocks listed on the New York Stock Exchange together with the S&P 500 index. Furthermore, we perform some analyses on the existence of a leverage effect in each of the series. In Section 5.3 we will present our estimation results for the SV model. Section 5.4 concludes.

5.1 The Mixture of Distributions Hypothesis and Stochastic Volatility Models

As was noted in the introduction many authors have tried to explain the sources of conditional heteroskedasticity in asset markets. One of the approaches that has been widely investigated is the Mixture of Distributions Hypothesis (MDH) that was originally developed by Clark (1973) and extended in Tauchen and Pitts (1983). Let y_t be the (continuously compounded) return on a financial asset. Suppose that this return is built from a number of independent increments δ_{it} :

$$y_t = \sum_{i=1}^{n_t} \delta_{it},$$

with n_t an integer random variable that determines the number of increments on day t . The δ_{it} are independently and identically distributed with mean zero and variance σ^2 . If we consider the increments δ_{it} as information events, n_t determines the rate at which information flows into the market. In this interpretation, n_t is a stochastic mixing variable. In the current setting n_t is the directing process for the subordinated process y_t . In other words, n_t governs the number of information arrivals δ_{it} , which in turn determine the daily return y_t . For example, if n_t is high, a lot of information flows to the market. Whether or not this leads to large price changes depends on the nature of the information events. Clark (1973) shows that when the number of information arrivals n_t is large a central limit theorem argument can be made that results in

$$y_t | n_t \sim N(0, \sigma^2 n_t). \quad (5.1)$$

Note that the conditioning variable n_t in (5.1) is generally unknown. Therefore, in empirical studies proxies for n_t have to be chosen. Typical proxies are the number of transactions in a day or the daily trading volume.

Lamoureux and Lastrapes (1990) use the framework in the previous paragraph to derive an ARCH model by assuming that the daily number of information arrivals is serially

correlated. They use the daily trading volume as a proxy for the rate of information arrival n_t . Similarly, we can make a link to Stochastic Volatility models. More specifically, rewrite (5.1) as

$$y_t = \sigma \sqrt{n_t} \epsilon_t, \quad \epsilon_t \sim N(0, 1).$$

In addition assume that the rate of information arrival follows the AR(1) process

$$n_t = \beta + \phi n_{t-1} + \eta_t, \quad \eta_t \sim N(0, \sigma_\eta^2).$$

Note that this is an SV model in which the volatility is represented by the number of information arrivals. As compared to a basic SV model (see Jacquier, Polson and Rossi, 1994) we need the additional assumption that n_t should be always positive. One way to ensure this is to let $h_t \equiv \ln n_t$ follow an AR(1) process. In this case we retrieve the lognormal SV model with AR(1) volatility:

$$y_t = \sigma \epsilon_t \exp\left(\frac{h_t}{2}\right), \quad \epsilon_t \sim N(0, 1), \quad (5.2)$$

$$h_t = \beta + \phi h_{t-1} + \eta_t, \quad \eta_t \sim N(0, \sigma_\eta^2). \quad (5.3)$$

From this we see that SV models are closely linked to the MDH. The variables h_t can either be interpreted as volatility or as a (transformed) information arrival process. Estimation of the SV model thus leads to an estimate of the information arrival process.

Lamoureux and Lastrapes (1990) test whether daily trading volume can be used as a proxy for the rate of information arrival by adding it additively to the GARCH equation. They find that the persistence of shocks to volatility decreases dramatically. It remains unclear, however, how their specification is related to the MDH. From (5.1) we see that the number of information arrivals n_t is multiplicatively related to the conditional variance of y_t . Therefore Bauer and Nieuwland (1994) suggest to introduce trading volume as a scale variable for the conditional volatility. This is also the line that we follow here.

The measurement equation (5.2) can be augmented to reflect the scale effects of volume on volatility in the following way

$$y_t = \sigma \epsilon_t \exp\left(\frac{h_t}{2}\right) V_t^{\lambda_y}, \quad \epsilon_t \sim N(0, 1). \quad (5.4)$$

In this equation V_t is a measure for daily trading volume. The parameter λ_y can be used to investigate how well the approximation is made. Transform (5.4) by squaring and taking logarithms. We get

$$\ln y_t^2 = h_t + 2\lambda_y \ln V_t + \beta_y + \xi_{y,t},$$

where $\beta_y \equiv \ln \sigma^2$ is a redefined constant, and $\xi_{y,t}$ has a log-chiquare distribution. If trading volume would be a good proxy for conditional volatility we expect λ_y to be significantly different from zero.

5.2 Data Description and Leverage Effects

The empirical analyses in this chapter are based on a sample of individual stocks listed on the New York Stock Exchange (NYSE). This sample was used in a recent study by Lamoureux and Lastrapes (1994).³ The period under investigation starts January 3, 1967, and ends December 30, 1987, spanning 5274 observations.⁴ The stocks we used are Alcoa, Amoco, Bristol-Myers, Champion and Hewlett-Packard. For each stock daily returns and daily trading volumes were collected for the specified period. Volume is measured as the number of shares traded in a day. Besides these five individual stocks we add daily returns of the S&P 500 composite stock index and the corresponding daily volumes of shares traded on the NYSE to our data set. The S&P 500 is a value-weighted, arithmetic, index of prices of common stock. The S&P 500 series is identical to the data used in Gallant, Rossi and Tauchen (1992), though we restrict the sample size to the same period as the individual stock series. Daily volume is measured as the total number of shares that are traded in a day for all stocks that make up the S&P 500. Returns and volumes of all series are adjusted for stock splits, stock- and cash dividends and capital modifications.

As in Gallant, Rossi and Tauchen (1992) we are mainly interested in modelling the short-run patterns in conditional heteroskedasticity and volumes. Therefore, we adjusted our data for systematic calendar effects, like weekends, holidays. In order to adjust both price and volume data, we perform an adjustment process in which systematic effects and deterministic trends are first removed from the mean and subsequently from the variance. In particular, we adjust for day-of-the-week, non-trading day and month-of-the-year effects. In addition we include linear and quadratic trends. This procedure is similar to the approach used in Gallant, Rossi and Tauchen (1992). Figures 5.1 and 5.2 reveal several notable facts. Figure 5.1 shows the unadjusted and adjusted return series. Note that the adjustment procedure does not seem to affect the series much. Inspection of Figure 5.2 reveals that the unadjusted trading volume series exhibits a clear trend in level. This deterministic trend is successfully removed in the lower panel of Figure 5.2.

Tables 5.1 and 5.2 provide summary statistics of the adjusted return series (y_t), the transformed return series ($\ln y_t^2$) and the adjusted log volume series ($\ln V_t$), respectively. The latter series are used as a proxy variable for volatility in the SV models that we will estimate. Table 5.1 shows that the adjusted return series feature slight negative skewness, consid-

³We are grateful to Christopher Lamoureux and William Lastrapes for providing the data set. The total data set contains data for ten stocks. For reasons of space, we decided to present results for five stocks from this set. Results for the remaining stocks do not differ substantially from the results presented.

⁴On Wednesdays in the last months of 1968, markets were closed due to paperwork backlog. We corrected for this phenomenon using non-trading day dummies.

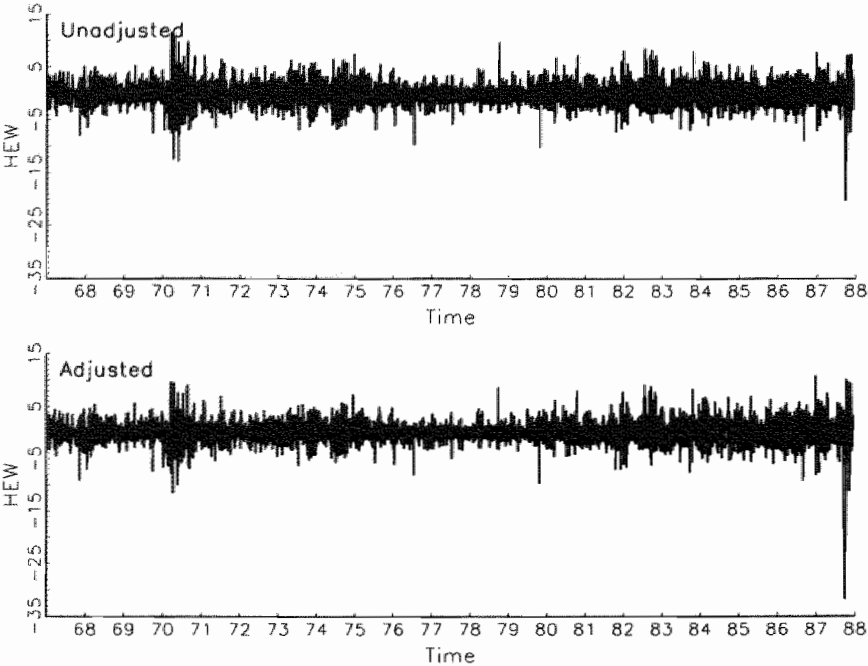


FIGURE 5.1: Hewlett-Packard: Returns

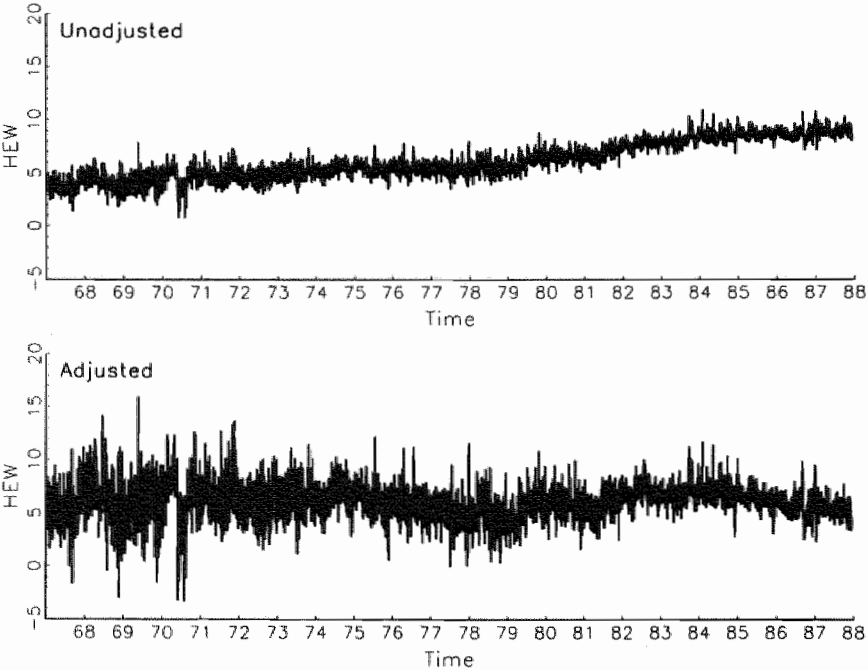


FIGURE 5.2: Hewlett-Packard: Trading Volumes (logarithms)

TABLE 5.1: Summary Statistics Adjusted Stock Returns

| | Adjusted Stock Returns: y_t | | | | | | Log Squared Adjusted Returns: $\ln y_t^2$ | | | | | |
|----------|-------------------------------|---------|---------|---------|---------|---------|---|---------|---------|---------|---------|---------|
| | ALC | AMO | BRI | CHA | HEW | S&P500 | ALC | AMO | BRI | CHA | HEW | S&P500 |
| Mean | 0.052 | 0.064 | 0.060 | 0.046 | 0.078 | 0.024 | -9.919 | -10.139 | -9.947 | -9.496 | -9.543 | -10.574 |
| St. dev. | 1.754 | 1.588 | 1.639 | 2.023 | 2.074 | 1.140 | 2.474 | 2.529 | 2.356 | 2.431 | 2.511 | 2.361 |
| Skewness | -0.535 | -0.578 | -0.396 | 0.008 | -0.716 | -0.513 | -1.082 | -1.248 | -1.125 | -1.133 | -1.358 | -1.265 |
| Kurtosis | 17.398 | 20.459 | 28.380 | 6.596 | 17.305 | 17.043 | 4.828 | 6.117 | 5.059 | 5.309 | 6.748 | 6.048 |
| Minimum | -25.386 | -26.564 | -27.806 | -18.403 | -31.677 | -18.025 | -23.926 | -26.629 | -22.748 | -25.894 | -28.070 | -27.453 |
| Maximum | 14.422 | 8.494 | 20.325 | 12.011 | 10.883 | 7.696 | -2.742 | -2.659 | -2.560 | -3.385 | -2.299 | -3.427 |
| ADF(4) | -37.041 | -34.118 | -33.225 | -32.694 | -32.353 | -31.593 | -26.052 | -23.991 | -27.101 | -26.944 | -26.113 | -26.806 |
| LB(25) | 195.5* | 104.7* | 125.1* | 65.9* | 57.3* | 223.4* | 620.0* | 1570.7* | 455.9* | 420.0* | 767.6* | 637.8* |
| LB(50) | 254.4* | 152.1* | 152.0* | 95.9* | 99.4* | 274.6* | 913.8* | 2532.3* | 696.2* | 533.2* | 1146.2* | 1025.7* |
| D(25) | 43.1 | 42.0 | 31.8 | 47.3* | 27.6 | 72.9* | 588.5* | 1483.6* | 445.3* | 375.2* | 760.2* | 637.3* |
| D(50) | 79.4* | 75.3 | 50.6 | 72.7 | 56.1 | 107.7* | 878.8* | 2437.6* | 681.3* | 491.3* | 1136.4* | 1013.9* |
| AC(1) | 0.163 | 0.110 | 0.070 | 0.074 | 0.059 | 0.193 | 0.107 | 0.167 | 0.127 | 0.105 | 0.139 | 0.078 |
| AC(2) | -0.018 | -0.034 | -0.059 | 0.005 | -0.025 | -0.014 | 0.087 | 0.151 | 0.102 | 0.078 | 0.106 | 0.057 |
| AC(3) | -0.042 | -0.026 | -0.058 | -0.006 | -0.013 | 0.003 | 0.076 | 0.121 | 0.066 | 0.072 | 0.075 | 0.043 |
| AC(4) | -0.029 | -0.033 | -0.040 | -0.017 | -0.010 | -0.010 | 0.084 | 0.102 | 0.050 | 0.083 | 0.079 | 0.066 |
| AC(5) | -0.016 | -0.026 | 0.044 | -0.009 | 0.014 | 0.002 | 0.089 | 0.139 | 0.075 | 0.061 | 0.089 | 0.093 |
| AC(10) | -0.013 | -0.002 | -0.052 | -0.005 | -0.001 | -0.017 | 0.063 | 0.108 | 0.055 | 0.061 | 0.087 | 0.051 |
| AC(20) | -0.012 | -0.021 | 0.010 | 0.002 | -0.026 | -0.010 | 0.040 | 0.100 | 0.051 | 0.049 | 0.069 | 0.066 |

Notes: ALC=Alcoa; AMO=Amoco; BRI=Bristol-Myers; CHA=Champion; HEW=Hewlett-Packard. Sample period: January 1967 - December 1987: 5274 observations. For adjustment procedure, see text. $LB(p)$ is the Ljung-Box statistic for serial correlation using p autocorrelations, $D(p)$ is the Diebold (1988) heteroskedasticity corrected serial correlation test, and $AC(p)$ denotes the sample autocorrelation at lag p . $ADF(4)$ is the augmented Dickey-Fuller test with 4 lags. The 1% MacKinnon critical value for rejection of the hypothesis of a unit root is -3.43.

TABLE 5.2: Summary Statistics Adjusted Log Volume

| | ALC | AMO | BRI | CHA | HEW | S&P500 |
|----------|---------|---------|----------|---------|----------|----------|
| Mean | 6.111 | 6.570 | 6.250 | 5.958 | 6.119 | 8.347 |
| St. dev. | 1.381 | 1.386 | 1.347 | 1.503 | 1.744 | 0.478 |
| Skewness | -0.183 | -0.328 | -0.091 | -0.161 | -0.271 | 0.233 |
| Kurtosis | 4.529 | 6.914 | 6.700 | 4.677 | 6.915 | 2.741 |
| Minimum | -3.760 | -5.296 | -3.691 | -3.565 | -11.227 | 6.173 |
| Maximum | 13.244 | 14.189 | 15.229 | 12.852 | 15.962 | 10.153 |
| ADF(4) | -20.560 | -16.033 | -17.721 | -19.211 | -18.890 | -10.106 |
| LB(25) | 5364.7* | 7645.3* | 9627.3* | 6342.9* | 8865.0* | 53019.3* |
| LB(50) | 7864.2* | 9534.7* | 12701.5* | 7844.3* | 12622.0* | 87710.8* |
| D(25) | 3988.6* | 4623.4* | 4986.6* | 4303.8* | 6010.5* | 34428.3* |
| D(50) | 4645.5* | 6778.7* | 7127.3* | 5592.6* | 8951.4* | 60729.3* |
| AC(1) | 0.383 | 0.544 | 0.452 | 0.371 | 0.447 | 0.837 |
| AC(2) | 0.309 | 0.456 | 0.364 | 0.322 | 0.379 | 0.759 |
| AC(3) | 0.260 | 0.416 | 0.342 | 0.294 | 0.328 | 0.734 |
| AC(4) | 0.249 | 0.420 | 0.325 | 0.264 | 0.312 | 0.717 |
| AC(5) | 0.217 | 0.390 | 0.317 | 0.260 | 0.275 | 0.703 |
| AC(10) | 0.184 | 0.332 | 0.273 | 0.250 | 0.264 | 0.640 |
| AC(20) | 0.136 | 0.313 | 0.217 | 0.155 | 0.226 | 0.577 |

Notes: See Table 5.1.

erable kurtosis and, measured by the Ljung-Box statistic, substantial serial correlation. Correcting the test statistic for heteroskedasticity, by using the Diebold statistic (Diebold, 1988), leads to somewhat lower values. Furthermore, all series are significantly first-order autocorrelated. The transformed return series are substantially negatively skewed, which is implied by the log transformation. The variance of each transformed series is higher than the variance of a log-chisquared ($\pi^2/2$) that arises in the standard stochastic volatility model. Table 5.2 presents the adjusted log volume series. The normality hypothesis can be rejected here too. In addition these series are persistently autocorrelated and serially correlated. Other features are again the significant serial correlation and persistent autocorrelation. For all series there appears to be no evidence of a unit root. In Appendix 5 we provide more descriptive analysis on our data through Vector AutoRegressive (VAR) analyses.

In the empirical literature on stock price behaviour some attention is given to the observation that positive and negative innovations to the return process have different impacts

on the conditional volatility. This so-called leverage effect was initially described by Black (1976), who found that negative shocks had a much bigger impact on conditional volatility than positive shocks. The explanation that Black gives is that a negative shock to a firm's stock price increases its leverage and thereby the expected variance of its stock. Glosten, Jagannathan and Runkle (1993) argue that the leverage effect is related to expected future cash flows. Assume that discount rates are constant. An unexpected negative shock in expected future cash flows will decrease the stock price. If the variance of future expected cash flows does not change proportionally, this variance per dollar of the stock price will rise. Consequently, future returns will be more volatile. Campbell and Hentschel (1992) develop a model in which they stress the link from conditional volatility to expected returns. Large pieces of news (either good or bad) on future dividends lead to falls in stock prices because they increase the required rate of return on the stocks. The drop in stock prices is damped in case of good news (higher future dividends) or amplified in case of bad news (lower future dividends). In contrast, small news events lead to lower volatility, lower required rate of return and higher stock prices. Campbell and Hentschel (1992) call the link between (conditional) volatility and required rates of return 'volatility feedback'. Although the causality between stock prices and volatility differs among the studies in this paragraph, they all stress the importance of correlation between volatility and stock returns.

Engle and Ng (1993) compare several models of the ARCH class with respect to their abilities to capture the leverage effect. The main tool they used is the news impact curve that plots the conditional volatility (h_t) against the innovation (ϵ_{t-1}) in the return series y_t . They also introduced some diagnostic tests based on the signs of a time series. The tests can be performed both on the residuals from a volatility model as well as on the original data y_t . In the latter case define $\epsilon_t \equiv y_t - m_y$ and $v_t \equiv \epsilon_t/s_y$, where m_y and s_y are the unconditional mean and standard deviation of y_t , respectively. Also define

$$\delta^-(\epsilon_{t-1}) = \begin{cases} 1 & \text{if } \epsilon_{t-1} < 0, \\ 0 & \text{otherwise,} \end{cases}$$

and $\delta^+(\epsilon_{t-1})$ analogously. The *Sign Bias Test* (SBT) is given by the t -statistic on α_1 of the least squares regression

$$\epsilon_t^2 = \alpha_0 + \alpha_1 \delta^-(\epsilon_{t-1}) + u_t.$$

This test examines the impact of negative and positive shocks to the squared innovations.⁵ The *Negative Sign Bias Test* (NSBT) and the *Positive Sign Bias Test* (PSBT) test whether the size of the innovations matters conditional on the innovations being negative or positive,

⁵Note that it does not matter if we have either ϵ_t^2 or v_t^2 as the dependent variable in the regression, because the constant scaling factor s_y in the definition of v_t does not influence the regression results.

TABLE 5.3: Sign Bias Tests

| | | ALC | AMO | BRI | CHA | HEW | S&P |
|-------------------------|------------|--------|--------|--------|--------|--------|--------|
| SBT | α_1 | 0.164 | 0.258 | 0.153 | 0.015 | 0.273 | 0.162 |
| | t | 0.494 | 0.862 | 0.405 | 0.056 | 0.573 | 1.215 |
| NSBT | α_1 | -1.894 | -1.725 | -2.875 | -0.654 | -1.556 | -1.265 |
| | t | -12.25 | -11.03 | -15.62 | -6.026 | -8.485 | -13.07 |
| PSBT | α_1 | 0.665 | 0.474 | 1.752 | 0.514 | 0.759 | 0.263 |
| | t | 2.249 | 3.137 | 9.312 | 4.885 | 3.985 | 2.668 |
| JSBT | LM | 231.9 | 177.0 | 465.6 | 92.94 | 123.8 | 250.5 |
| Volume corrected series | | | | | | | |
| SBT | α_1 | 0.188 | 0.351 | 0.189 | 0.163 | -0.314 | 0.154 |
| | t | 0.562 | 1.163 | 0.499 | 0.619 | 0.658 | 1.117 |
| NSBT | α_1 | -1.932 | -1.722 | -2.858 | -0.667 | -1.561 | -1.266 |
| | t | -12.51 | -11.09 | -15.54 | -6.194 | -8.531 | -12.75 |
| PSBT | α_1 | 0.656 | 0.445 | 1.755 | 0.468 | 0.752 | 0.225 |
| | t | 4.120 | 2.889 | 9.261 | 4.358 | 3.931 | 2.164 |
| JSBT | LM | 237.7 | 173.8 | 459.7 | 87.65 | 123.8 | 232.04 |

Notes: Results are based on residuals of an AR(15) (upper part) and an AR(15) plus lagged volume (lower part). SBT is the *Sign Bias Test*, NSBT the *Negative Sign Bias Test*, PSBT the *Positive Sign Bias Test*, and JSBT the *Joint Sign Bias Test*. See text for description of regression equations. Reported are the coefficients on the sign variables together with the t -statistic (SBT, NSBT, PSBT) and the LM test statistic ($\chi^2(3)$ distributed) for JSBT. See Table 5.1 for description of the series.

respectively. The regressions are

$$\text{NSBT:} \quad \epsilon_t^2 = \alpha_0 + \alpha_1 \delta^-(\epsilon_{t-1}) \epsilon_{t-1} + u_t,$$

$$\text{PSBT:} \quad \epsilon_t^2 = \alpha_0 + \alpha_1 \delta^+(\epsilon_{t-1}) \epsilon_{t-1} + u_t.$$

Engle and Ng (1993) also report a *Joint Sign Bias Test* (JSBT) in which all three individual tests are combined. It is an LM test for adding three extra variables under a model for the conditional volatility. In case of a diagnostic test for the raw data this test is equal to a standard F -test. The regression for the JSBT is

$$\epsilon_t^2 = \alpha_0 + \alpha_{11} \delta^-(\epsilon_{t-1}) + \alpha_{12} \delta^-(\epsilon_{t-1}) \epsilon_{t-1} + \alpha_{13} \delta^+(\epsilon_{t-1}) \epsilon_{t-1} + u_t.$$

The test statistic is given by T times the R -squared from this regression and is asymptotically chi-squared distributed with 3 degrees of freedom. Results are reported in Table 5.3 for the raw data. We use the residuals from an autoregression of order 15 in order

to delete possible serial dependence in y_t . There appears to be no strong evidence for an asymmetric effect of innovations to volatility as is shown by the insignificant coefficients α_1 in the SBT. The NSBT and PSBT results show that a size effect seems to be present: large negative/positive shocks have more effect on volatility than smaller shocks. These results remain qualitatively the same when we include the logarithm of lagged volume ($\ln V_{t-1}$) as an extra regressor in the autoregression.

Although the leverage effect is considered to be an intrinsic characteristic of stock prices, Gallant, Rossi and Tauchen (1992) have shown that the leverage effect arises predominantly as a tail phenomenon in the distribution of the S&P 500 index series from 1928 to 1987. When large negative (and positive) shocks are deleted from the sample, the evidence for existence of the leverage effect vanishes quickly. This can be tentatively shown by doing the regressions

$$|y_t| = \alpha_0 + \alpha_1 |y_{t-1}| + \beta [y_{t-1} \delta^-(y_{t-1})] + u_t. \quad (5.5)$$

We use the absolute value of returns $|y_t|$ for describing volatility. Thus, an autoregressive process of order 1 is fitted to the data in which negative returns are treated differently than positive returns through the term $y_{t-1} \delta^-(y_{t-1})$. The coefficient β captures the leverage effect. The null hypothesis of no leverage is that $\beta=0$. In the upper part of Table 5.4 we present results from doing the regressions (5.5) for the five stock return series and the S&P 500 index. The cut-off points in the first column are determined from the unconditional standard deviations s_y of the y_t series. Thus, at a particular c we restrict our sample to those points (y_t, y_{t-1}) that obey the restrictions $|y_t| < s_y c$ and $|y_{t-1}| < s_y c$. When $c = \infty$, the whole sample y_t ($t = 1, \dots, T$) is used. From the table we see that the coefficients β are generally insignificant for lower cutoff points. Furthermore, a significant leverage effect can only be found in three series (Alcoa, Amoco and the S&P 500). In case of Alcoa only two observations (related to the October 1987 crash) cause β to become significantly negative. Results for the S&P 500 index series suggest that this series seems to have a significant leverage effect.

Gallant, Rossi and Tauchen (1992) also provide evidence for the disappearance of the leverage effect once volume is taken into account. They perform a similar regression as (5.5) in which the asymmetry coefficient is now a linear function of volume:

$$|y_t| = \alpha_0 + \alpha_1 |y_{t-1}| + (\gamma_0 + \gamma_1 \ln V_{t-1}) y_{t-1} \delta^-(y_{t-1}) + u_t.$$

The lower part of Table 5.4 shows the results from this regression. The coefficients $\beta = \hat{\gamma}_0 + \hat{\gamma}_1 \ln V_{t-1}$ are evaluated at the volumes $\ln V_{t-1} = m_V + k s_V$, ($k = -2, -1, 0, 1, 2$), where m_V and s_V are the unconditional mean and unconditional standard deviation of the $\ln V_{t-1}$ series, respectively. From the table we see that volume interacts with β . When $k=0$ only the coefficient for the S&P 500 is still significant, but has a much lower absolute value.

TABLE 5.4: Leverage Regressions

| c | | ALC | AMO | BRI | CHA | HEW | S&P500 |
|---|------------|--------|--------|--------|--------|--------|--------|
| 1 | β | 0.015 | 0.014 | 0.025 | -0.019 | -0.007 | -0.007 |
| | $s(\beta)$ | 0.019 | 0.021 | 0.019 | 0.019 | 0.020 | 0.019 |
| | T | 4888 | 4821 | 4897 | 4837 | 4843 | 4850 |
| 2 | β | 0.004 | -0.048 | 0.022 | -0.017 | -0.004 | -0.068 |
| | $s(\beta)$ | 0.020 | 0.020 | 0.018 | 0.019 | 0.019 | 0.018 |
| | T | 5229 | 5229 | 5233 | 5232 | 5228 | 5222 |
| 3 | β | 0.024 | -0.049 | 0.032 | 0.000 | 0.021 | -0.067 |
| | $s(\beta)$ | 0.020 | 0.020 | 0.018 | 0.019 | 0.019 | 0.018 |
| | T | 5263 | 5265 | 5260 | 5267 | 5269 | 5266 |
| 4 | β | 0.036 | -0.057 | 0.002 | -0.001 | 0.000 | -0.041 |
| | $s(\beta)$ | 0.020 | 0.020 | 0.020 | 0.019 | 0.019 | 0.018 |
| | T | 5269 | 5271 | 5269 | 5273 | 5270 | 5272 |
| 5 | β | 0.021 | -0.049 | -0.006 | -0.001 | 0.004 | -0.064 |
| | $s(\beta)$ | 0.020 | 0.019 | 0.019 | 0.019 | 0.019 | 0.019 |
| | T | 5271 | 5273 | 5270 | 5273 | 5273 | 5273 |
| ∞ | β | -0.042 | -0.049 | 0.003 | -0.001 | 0.004 | -0.064 |
| | $s(\beta)$ | 0.019 | 0.019 | 0.019 | 0.019 | 0.019 | 0.019 |
| | T | 5273 | 5273 | 5273 | 5273 | 5273 | 5273 |
| Asymmetry coefficients for various values of volume | | | | | | | |
| k | | | | | | | |
| -2 | β | 0.086 | 0.069 | 0.080 | -0.070 | -0.014 | -0.019 |
| | $s(\beta)$ | 0.036 | 0.042 | 0.036 | 0.035 | 0.040 | 0.029 |
| -1 | β | 0.031 | 0.023 | 0.050 | -0.039 | -0.007 | -0.038 |
| | $s(\beta)$ | 0.026 | 0.030 | 0.026 | 0.025 | 0.029 | 0.023 |
| 0 | β | -0.026 | -0.023 | 0.020 | -0.008 | 0.000 | -0.058 |
| | $s(\beta)$ | 0.020 | 0.021 | 0.020 | 0.020 | 0.021 | 0.019 |
| 1 | β | -0.081 | -0.068 | -0.010 | 0.022 | 0.007 | -0.077 |
| | $s(\beta)$ | 0.021 | 0.020 | 0.020 | 0.021 | 0.020 | 0.020 |
| 2 | β | -0.137 | -0.114 | -0.040 | 0.053 | 0.014 | -0.097 |
| | $s(\beta)$ | 0.029 | 0.028 | 0.026 | 0.030 | 0.027 | 0.025 |

Notes: Upper part shows the asymmetry coefficient and its standard error in the least squares regression $|y_t| = \alpha_0 + \alpha_1|y_{t-1}| + \beta[y_{t-1}\delta^-(y_{t-1})] + u_t$, which is truncated to points (y_t, y_{t-1}) that have both elements to be smaller than cs_y in absolute value. T is the number of observations used in the regression. Lower part shows asymmetry coefficients $\beta = \hat{\gamma}_0 + \hat{\gamma}_1 \ln V_{t-1}$ from the regression $|y_t| = \alpha_0 + \alpha_1|y_{t-1}| + (\gamma_0 + \gamma_1 \ln V_{t-1}) y_{t-1} \delta^-(y_{t-1}) + u_t$ evaluated at volumes $m_V + ks_V$, with m_V and s_V the unconditional mean and standard deviation of $\ln V_{t-1}$. See Table 5.1 for description of series.

For this series we find that increasing volume leads to a stronger leverage effect. For the other series we sometimes find a positive leverage effect, meaning that good news has more impact on volatility than negative news.

In conclusion, the exercises above show that evidence for a leverage effect is rather mixed. Besides Gallant, Rossi and Tauchen (1992), Engle and Lee (1993) provide evidence that there is no leverage effect in the long run in a model where conditional volatility is decomposed into a permanent and a transitory component. In our data series, it seems that there is a significant leverage effect for the S&P 500 index, but not for individual stock series. Furthermore, the interactions with volume are not clear from the regression results. As Gallant, Rossi and Tauchen (1992) indicate, a parametric model that jointly describes stock returns and volumes needs to possess some nonlinear features. Nevertheless, in some of the models we estimate below we will allow the possibility of a link between the conditional variance and innovations to the return process. It remains an open question, though, to what extent results from this link can be interpreted as a leverage effect.

5.3 Estimation Results

As was mentioned in Section 5.1 we propose to use volume as a scale factor for the volatility. We chose to model time-varying volatility by the SV model, where the variance is modeled as an unobserved component that behaves stochastically. Preliminary evidence suggests that SV models capture the dynamics of the conditional volatility in a flexible way (Jacquier, Polson and Rossi, 1994). The state-space specification we use, has been introduced by Harvey and Shephard (1994). This specification incorporates a leverage effect by allowing correlation between measurement and transition equation disturbances.

We restate the SV model that was derived in Section 5.1 here, with volume as a scale factor

$$y_t = \sigma \epsilon_t \exp\left(\frac{h_t}{2}\right) V_t^{\lambda_v}, \quad \epsilon_t \sim N(0, 1), \quad (5.6)$$

$$h_{t+1} = \beta + \phi h_t + \eta_t, \quad \eta_t \sim N(0, \sigma_\eta^2). \quad (5.7)$$

The innovations in the measurement and transition equation are contemporaneously correlated: $\text{Corr}(\epsilon_t, \eta_t) = \rho$. We furthermore assume that $\text{Corr}(\epsilon_t, \eta_{t-k}) = 0$, for $k \neq 0$. The h_t process is assumed to be AR(1), but other specifications are possible. A negative contemporaneous correlation can be interpreted as the usual leverage effect: negative shocks in stock prices increase volatility more than positive shocks. Note that we shifted the condi-

TABLE 5.5: Stochastic Volatility Estimation Results I:

$$\ln y_t^2 = h_t + \beta_y + \xi_{y,t}, \quad \xi_y \sim N(0, \pi^2/2),$$

$$h_{t+1} = \phi h_t + \eta_t, \quad \eta_t \sim N(0, \sigma_\eta^2)$$

| | ALC | AMO | BRI | CHA | HEW | S&P |
|---------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|
| β_y | -0.666 (0.358) | -0.865 (1.995) | -0.757 (0.254) | -0.303 (0.136) | -0.316 (0.365) | -1.335 (0.642) |
| ϕ | 0.982 (0.003) | 0.995 (0.007) | 0.973 (0.002) | 0.916 (0.004) | 0.974 (0.002) | 0.992 (0.004) |
| σ_η | 0.130 (0.004) | 0.082 (0.004) | 0.163 (0.006) | 0.340 (0.014) | 0.190 (0.006) | 0.085 (0.002) |
| LogL | -12017 | -11951 | -12070 | -12264 | -12237 | -11691 |
| Min | -8.409 | -5.557 | -6.555 | -5.831 | -8.183 | -5.950 |
| Max | 2.712 | 3.110 | 2.853 | 2.481 | 2.688 | 3.096 |
| Skew | -1.264 | -1.221 | -1.365 | -1.280 | -1.542 | -1.286 |
| Kurt | 6.155 | 5.430 | 6.078 | 5.451 | 7.624 | 5.726 |
| LB(10) | 5.340 | 19.677 | 25.070 | 15.045 | 19.153 | 16.770 |
| LB(20) | 16.370 | 27.261 | 38.373 | 23.744 | 38.647 | 28.785 |
| SBT | 0.865 | -0.678 | 0.968 | 1.529 | 1.253 | 1.068 |
| NSBT | -1.507 | 1.694 | -1.813 | -1.800 | 0.098 | -0.052 |
| PSBT | 0.009 | -0.102 | -1.024 | -1.795 | -0.840 | -0.191 |
| JSBT | 3.066 | 3.834 | 3.552 | 4.511 | 2.684 | 2.468 |

Notes: Heteroskedasticity-consistent standard errors are given in parentheses. *Min*, *Max*, *Skew*, and *Kurt* are the minimum, maximum, skewness, and kurtosis, respectively, of the standardized one-step prediction errors. *LB(.)* are Ljung-Box statistics, and *SBT*, *NSBT*, *PSBT*, and *JSBT* are sign bias tests (see text). See Table 5.1 for description of series.

tional volatility one period forward in (5.7). In this way a shock in ϵ_t leads to a shock in next periods volatility h_{t+1} through η_t .

The model above can be linearized by squaring the observations and taking logarithms. The model we use in the empirical part can be written in a linear state-space format as follows

$$\ln y_t^2 = h_t + \lambda_y \ln V_t + \beta_y + \xi_{y,t}, \quad (5.8)$$

$$h_{t+1} = \phi h_t + \eta_t. \quad (5.9)$$

The constant parameter β_y incorporates both the constants σ from (5.6) and β from (5.7), without loss of generality. In the estimation of the state-space system (5.8) and (5.9) we have to take account of the correlation structure between the measurement and the state equation.

TABLE 5.6: Stochastic Volatility Estimation Results II:

$$\ln y_t^2 = h_t + \beta_y + \xi_{y,t}, \quad \xi_y \sim N(0, \pi^2/2),$$

$$h_{t+1} = (\phi - 0.2242\rho\sigma_\eta s_t)h_t + s_t\sigma_\eta\rho(0.7979 + 0.2242\ln y_t^2) + \eta_t^+,$$

$$\eta_t^+ \sim N(0, \sigma_\eta^2(1 - 0.8846\rho^2))$$

| | ALC | AMO | BRI | CHA | HEW | S&P500 |
|---------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|
| β_y | -0.661 (0.350) | -0.879 (1.898) | -0.781 (0.282) | -0.307 (0.144) | -0.327 (0.399) | -1.313 (0.393) |
| ϕ | 0.981 (0.005) | 0.995 (0.001) | 0.977 (0.006) | 0.924 (0.033) | 0.977 (0.006) | 0.986 (0.003) |
| σ_η | 0.132 (0.020) | 0.081 (0.010) | 0.152 (0.023) | 0.322 (0.093) | 0.176 (0.029) | 0.120 (0.018) |
| ρ | 0.070 (0.121) | -0.023 (0.260) | -0.307 (0.109) | -0.096 (0.067) | -0.113 (0.092) | -0.481 (0.149) |
| LogL | -12017 | -11951 | -12065 | -12263 | -12236 | -11680 |
| Min | -8.404 | -5.547 | -6.619 | -5.861 | -8.213 | -5.961 |
| Max | 2.720 | 3.109 | 2.874 | 2.454 | 2.680 | 2.970 |
| Skew | -1.264 | -1.221 | -1.368 | -1.283 | -1.544 | -1.305 |
| Kurt | 6.159 | 5.430 | 6.092 | 5.461 | 7.642 | 5.770 |
| LB(10) | 5.209 | 19.990 | 24.977 | 14.770 | 18.391 | 19.398 |
| LB(20) | 15.768 | 27.517 | 38.373 | 23.014 | 38.503 | 31.033 |
| SBT | 0.712 | -0.704 | 0.938 | 1.356 | 1.241 | 1.126 |
| NSBT | -1.541 | 1.690 | -1.692 | -1.762 | 0.089 | -0.210 |
| PSBT | 0.013 | -0.099 | -0.931 | -1.751 | -0.796 | -0.366 |
| JSBT | 3.016 | 3.827 | 3.019 | 4.460 | 2.615 | 2.072 |

Notes: Heteroskedasticity-consistent standard errors are given in parentheses. *Min*, *Max*, *Skew*, and *Kurt* are the minimum, maximum, skewness, and kurtosis, respectively, of the standardized one-step prediction errors. *LB(.)* are Ljung-Box statistics, and *SBT*, *NSBT*, *PSBT*, and *JSBT* are sign bias tests (see text). See Table 5.1 for description of series.

The estimation of SV models has received considerable attention lately. A number of different approaches have been proposed. Formally, Maximum Likelihood (ML) estimation requires a method to find the solution of a high-dimensional integral, that arises due to the latent character of the volatility equation and the non-normality of $\xi_{y,t}$. One way to calculate the likelihood is to use simulations. For any details about this procedure, see Jacquier, Polson and Rossi (1994). Another approach is to approximate the error distribution of $\xi_{y,t}$ by a normal distribution and apply the usual Kalman filtering techniques. Estimating SV models in this way is called Quasi Maximum Likelihood (QML) estimation. This method implies that only the first two moments of the log chisquare distribution are needed. From Abramowitz and Stegun (1970) we find that $E\xi = -1.2704$ and $\text{Var}(\xi) = \pi^2/2$.⁶

⁶Ruiz (1994) shows that QML parameter estimates become more efficient, compared to parameter

TABLE 5.7: Stochastic Volatility Estimation Results III:

$$\ln y_t^2 = h_t + \beta_y + \lambda_y \ln V_t + \xi_{y,t}, \quad \xi_y \sim N(0, \pi^2/2),$$

$$h_{t+1} = \phi h_t + \eta_t, \quad \eta_t \sim N(0, \sigma_\eta^2)$$

| | ALC | AMO | BRI | CHA | HEW | S&P500 |
|---------------|-------------------|-------------------|-------------------|-------------------|-------------------|--------------------|
| β_y | -1.594 (0.396) | -3.263 (1.650) | -2.814 (0.505) | -1.980 (0.275) | -2.215 (0.587) | -11.014 (1.997) |
| λ_y | 0.152 (0.029) | 0.366 (0.035) | 0.331 (0.033) | 0.282 (0.029) | 0.312 (0.027) | 1.160 (0.153) |
| ϕ | 0.983 (0.004) | 0.995 (0.002) | 0.985 (0.002) | 0.957 (0.008) | 0.989 (0.003) | 0.994 (0.001) |
| σ_η | 0.125 (0.018) | 0.077 (0.008) | 0.125 (0.010) | 0.229 (0.028) | 0.104 (0.015) | 0.092 (0.004) |
| LogL | -11998 | -11855 | -11992 | -12193 | -12121 | -11626 |
| Min | -8.500 | -5.545 | -6.438 | -6.147 | -8.322 | -5.984 |
| Max | 2.578 | 2.726 | 2.330 | 2.086 | 2.458 | 2.162 |
| Skew | -1.284 | -1.295 | -1.423 | -1.324 | -1.610 | -1.338 |
| Kurt | 6.239 | 5.667 | 6.210 | 5.571 | 7.834 | 5.798 |
| LB(10) | 5.433 | 11.900 | 18.845 | 7.997 | 10.572 | 17.127 |
| LB(20) | 18.007 | 19.199 | 33.877 | 18.091 | 29.227 | 25.599 |
| SBT | 0.492 | 0.478 | 1.627 | 2.496 | 0.306 | 1.884 |
| NSBT | -1.627 | 1.615 | -1.769 | -1.132 | 0.024 | -0.023 |
| PSBT | -0.014 | 0.064 | -1.509 | -1.616 | -0.167 | -0.818 |
| JSBT | 3.322 | 6.612 | 3.840 | 6.493 | 0.167 | 6.410 |

Notes: Heteroskedasticity-consistent standard errors are given in parentheses. *Min*, *Max*, *Skew*, and *Kurt* are the minimum, maximum, skewness, and kurtosis, respectively, of the standardized one-step prediction errors. *LB(.)* are Ljung-Box statistics, and *SBT*, *NSBT*, *PSBT*, and *JSBT* are sign bias tests (see text). See Table 5.1 for description of series.

The state-space model (5.8) and (5.9) cannot be estimated via the Kalman filter directly, since inference about the correlation parameter is invalid. This problem arises since by squaring the observations information about the sign of the correlation coefficient is lost. Harvey and Shephard (1994) (HS94 hereafter) develop a method to estimate ρ within the state-space formulation. It is based on the observation that the distribution of the signs is independent of the parameters in the model. In other words the signs of the observations are ancillary statistics. This means that we can condition on the signs in order to find an estimate for ρ . We follow the HS94 approach and find the following linear state space model

$$\ln y_t^2 = h_t + \lambda_y \ln V_t + \beta_y + \xi_{y,t},$$

estimates from Generalized Method of Moments (GMM), when the number of observations increases.

TABLE 5.8: Stochastic Volatility Estimation Results IV:

$$\ln y_t^2 = h_t + \beta_y + \lambda_y \ln V_t + \xi_{y,t}, \quad \xi_y \sim N(0, \pi^2/2),$$

$$h_{t+1} = (\phi - 0.2242\rho\sigma_\eta s_t)h_t + s_t\sigma_\eta\rho(0.7979 + 0.2242\ln y_t^2) + \eta_t^+,$$

$$\eta_t^+ \sim N(0, \sigma_\eta^2(1 - 0.8846\rho^2))$$

| | ALC | AMO | BRI | CHA | HEW | S&P500 |
|---------------|-------------------|-------------------|-------------------|-------------------|-------------------|--------------------|
| β_y | -1.595 (0.399) | -3.306 (1.731) | -2.848 (0.510) | -2.010 (0.288) | -2.209 (0.650) | -11.506 (1.285) |
| λ_y | 0.152 (0.029) | 0.366 (0.035) | 0.331 (0.033) | 0.285 (0.029) | 0.311 (0.027) | 1.228 (0.155) |
| ϕ | 0.983 (0.004) | 0.996 (0.001) | 0.986 (0.002) | 0.962 (0.007) | 0.990 (0.003) | 0.985 (0.005) |
| σ_η | 0.125 (0.018) | 0.074 (0.006) | 0.125 (0.011) | 0.213 (0.027) | 0.096 (0.014) | 0.127 (0.021) |
| ρ | 0.015 (0.128) | -0.136 (0.303) | -0.338 (0.135) | -0.199 (0.082) | -0.185 (0.116) | -0.514 (0.124) |
| LogL | -11998 | -11855 | -11987 | -12190 | -12120 | -11604 |
| Min | -8.499 | -5.565 | -6.493 | -5.898 | -8.350 | -5.972 |
| Max | 2.579 | 2.706 | 2.345 | 2.390 | 2.455 | 1.979 |
| Skew | -1.284 | -1.296 | -1.426 | -1.330 | -1.613 | -1.365 |
| Kurt | 6.240 | 5.667 | 6.226 | 5.597 | 7.853 | 5.881 |
| LB(10) | 5.431 | 12.326 | 18.660 | 7.101 | 10.492 | 18.855 |
| LB(20) | 17.985 | 19.658 | 33.627 | 17.994 | 28.762 | 28.374 |
| SBT | 0.509 | 0.575 | 1.375 | 2.212 | 0.564 | 1.608 |
| NSBT | -1.633 | 1.623 | -1.643 | -1.483 | 0.015 | -0.283 |
| PSBT | -0.013 | 0.094 | -1.489 | -1.382 | -0.171 | -0.944 |
| JSBT | 3.341 | 7.359 | 3.436 | 4.994 | 0.625 | 3.563 |

Notes: Heteroskedasticity-consistent standard errors are given in parentheses. *Min*, *Max*, *Skew*, and *Kurt* are the minimum, maximum, skewness, and kurtosis, respectively, of the standardized one-step prediction errors. *LB(.)* are Ljung-Box statistics, and *SBT*, *NSBT*, *PSBT*, and *JSBT* are sign bias tests (see text). See Table 5.1 for description of series.

$$h_{t+1} = \phi h_t + s_t \kappa + \eta_t^*,$$

$$\begin{pmatrix} \xi_{y,t} \\ \eta_t^* \end{pmatrix} | s_t \sim IID \left(0, \begin{bmatrix} \sigma_\xi^2 & \nu s_t \\ \nu s_t & \sigma_\eta^2 - \kappa^2 \end{bmatrix} \right),$$

where $\nu = \text{Cov}_+(\eta_t, \xi_t)$, i.e. ν is a conditional covariance where we condition on positive observations. Note that the measurement equation does not change. The new transition equation now depends on the sign of the observations

$$s_t = \begin{cases} 1 & \text{if } y_t > 0, \\ -1 & \text{if } y_t \leq 0. \end{cases}$$

Define κ to be the expectation of η_t conditional on s_t :

$$\kappa = E(\eta_t | s_t = 1) = -E(\eta_t | s_t = -1).$$

The last equivalence follows since the joint distribution of ϵ_t and η_t is assumed to be symmetric. See Harvey, Ruiz and Shephard (1994) for more details. Using the ancillarity of s_t and the results in Harvey (1989) we obtain the following state-space model with uncorrelated disturbances

$$\begin{aligned} \ln y_t^2 &= h_t + \lambda_y \ln V_t + \beta_y + \xi_{y,t}, \\ h_{t+1} &= \left(\phi - \frac{\nu s_t}{\sigma_\xi^2} \right) h_t + s_t \left(\kappa + \frac{\nu}{\sigma_\xi^2} \{ \ln y_t^2 - \lambda_y \ln V_t - \beta_y \} \right) + \eta_t^+, \\ \begin{pmatrix} \xi_{y,t} \\ \eta_t^+ \end{pmatrix} | s_t &\sim IID \left(0, \begin{bmatrix} \sigma_\xi^2 & 0 \\ 0 & \sigma_\eta^2 - \kappa^2 - \frac{\nu^2}{\sigma_\xi^2} \end{bmatrix} \right). \end{aligned} \quad (5.10)$$

Like HS94 we assume that ϵ_t and η_t are bivariate normal with correlation parameter ρ . This means that

$$\kappa = E(\eta_t | s_t = 1) = -E(\eta_t | s_t = -1) = 0.7979 \rho \sigma_\eta,$$

$$\nu = Cov(\eta_t, \xi_t | s_t = 1) = -Cov(\eta_t, \xi_t | s_t = -1) = 1.1061 \rho \sigma_\eta.$$

Inserting these values in (5.10) yields the linear state space model used in the estimation. For convenience, we set σ_ξ^2 equal to $\pi^2/2$ in the QML specifications.

In tables 5.5 through 5.8 we present the QML stochastic volatility results for our final specification. Like HS94 we estimated all models on the residuals of an AR(1) of y_t in order to remove any first-order serial correlation from the return series. Table 5.5 provides parameter estimates for a model without volume and without leverage. Results are in concordance with typical results from financial time series estimated in a SV framework. A high persistence of volatility ϕ , and a low signal-noise ratio ($\sigma_\eta^2/\sigma_\xi^2$). Including the leverage effect through ρ in Table 5.6 leads to mixed results. Only two series, including the S&P 500, exhibit a significant leverage effect. Tables 5.7 and 5.8 present results after including volume in the measurement equation. Note that in both tables λ_y is significant for all series. Significant leverage effects are found in three cases now, but changes in the leverage parameter between Table 5.6 and Table 5.8 are not of same sign, *i.e.* including trading volume does not influence leverage parameters in a uniform way. Volume enhances the fit, measured by the (log) likelihood, of the basic SV model significantly. In every case likelihood ratio statistics are larger than the critical value 6.63 at a 1% significance level.

We also computed some diagnostic test statistics using one-step ahead standardized prediction errors from the models. The Ljung-Box test statistics show that for some series the

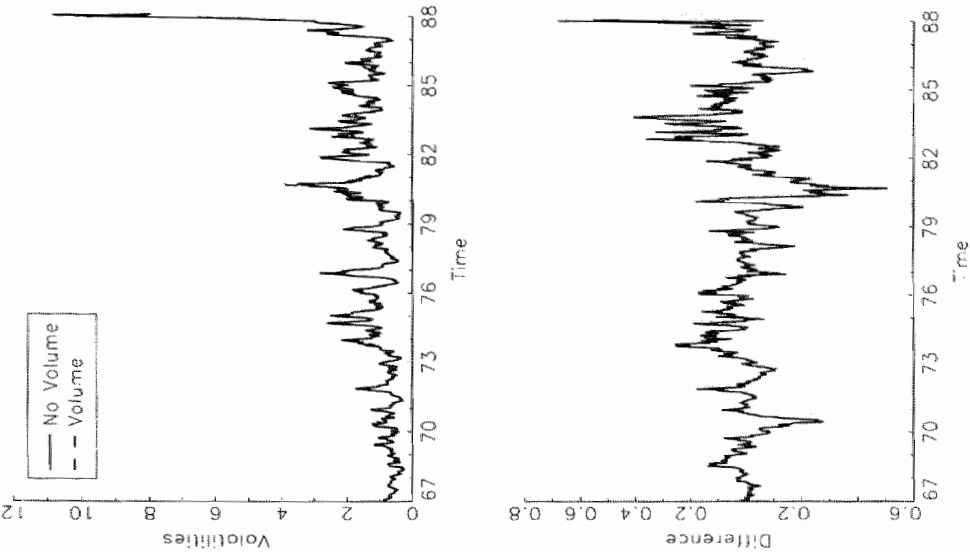


FIGURE 5.3: Volatilities Alcoa

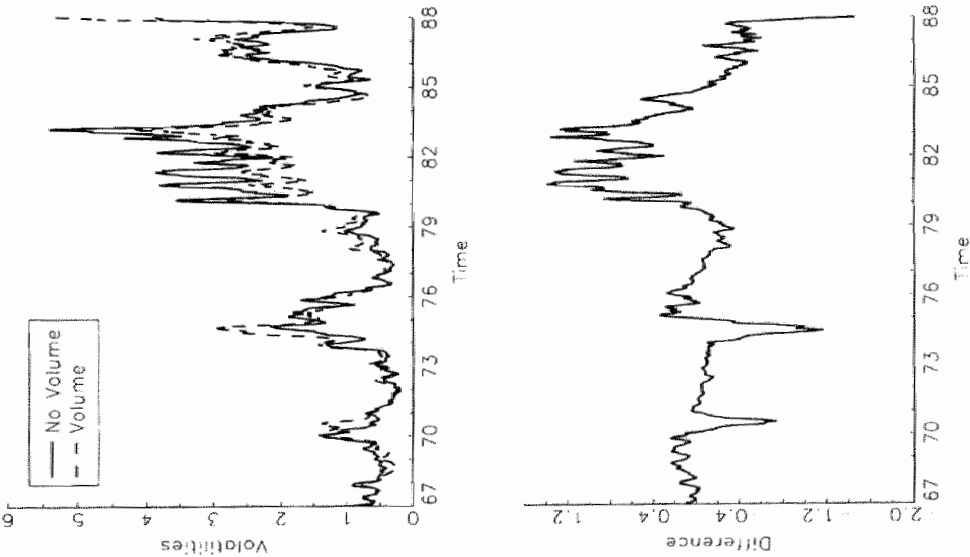


FIGURE 5.4: Volatilities Amoco

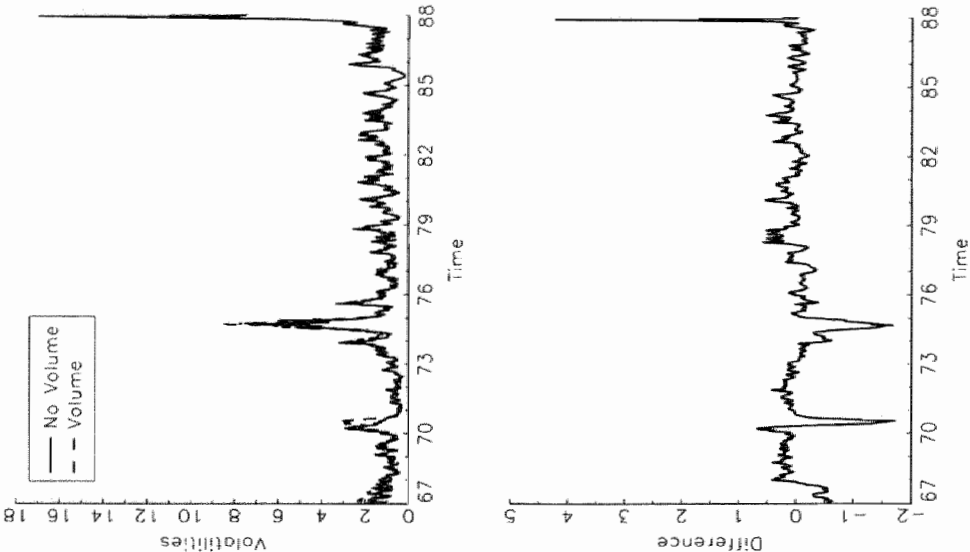


FIGURE 5.5: Volatilities Bristol-Myers

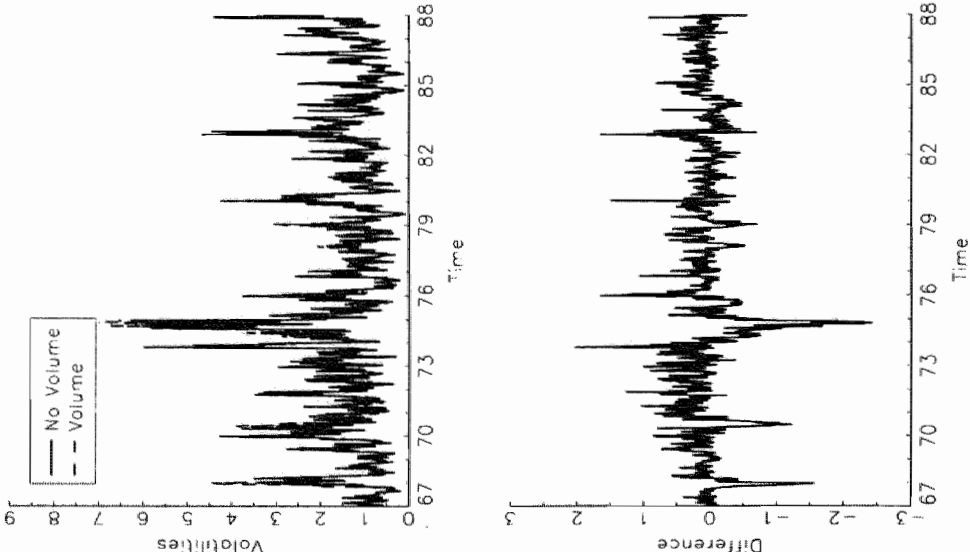


FIGURE 5.6: Volatilities Champion

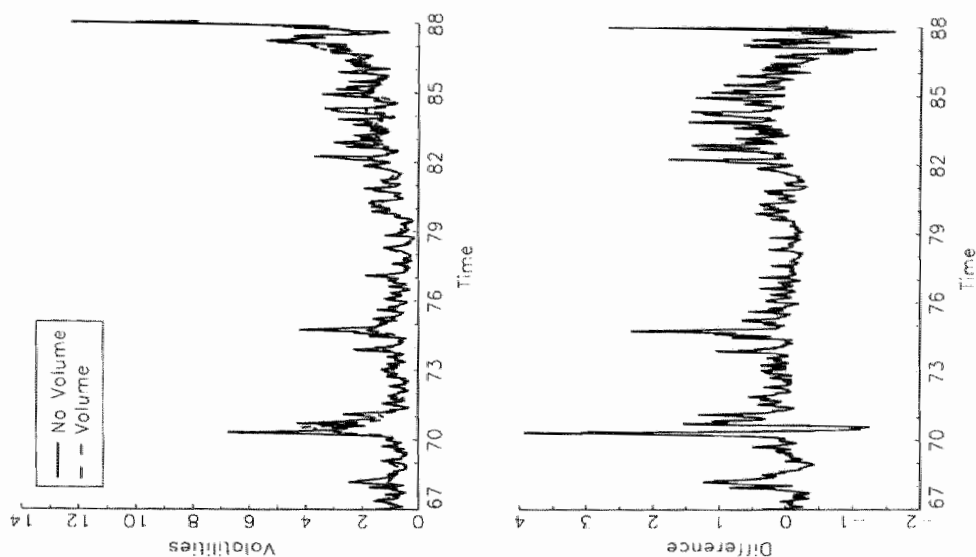


FIGURE 5.7: Volatilities Hewlett-Packard

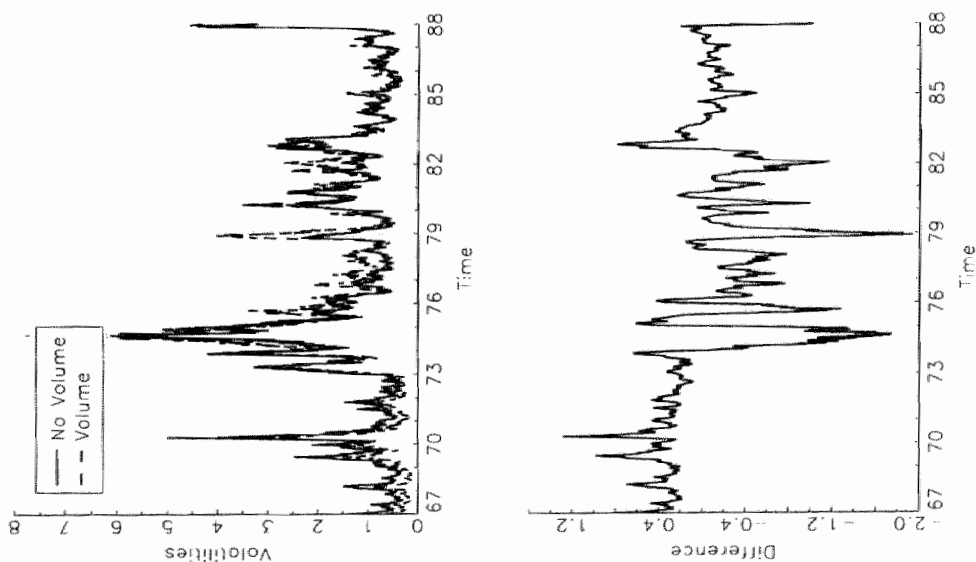


FIGURE 5.8: Volatilities S&P 500

AR(1) dynamic specification for the conditional variance is not sufficient. Furthermore, kurtosis remains high. The sign bias tests show that sign effects are in general not present in the one-step-ahead prediction errors of the Kalman filter.

In figures 5.3 to 5.8 we present smoothed volatility series for the SV models in tables 5.6 and 5.8. Let $h_{i|T}$ be the smoothed logarithmic volatility and $P_{i|T}$ the corresponding smoothed variance. If we assume that the volatility series are lognormally distributed, the volatility can be computed as $\exp(h_{i|T} + \frac{1}{2}P_{i|T})$. Besides the two volatility series we also plotted the difference between these two series, which is computed as the volatility from a benchmark model without volume minus the volatility of the model that includes volume in the measurement equation. The units on the vertical axes should be read as percentage squared per day. The volatility series differ markedly for the stocks in our sample. For Alcoa, Amoco, Bristol-Myers and Hewlett-Packard volatility seems to increase in the latter part of the sample. This cannot be said for Champion and, interestingly, the S&P 500 index series. In all series the stock market crash in October 1987 leads to higher volatility. Introducing trading volume in the SV model sometimes leads to persistent deviations of volatility from the benchmark volatility. This is especially the case for Amoco in the beginning of the 1980s and the S&P 500 index in the first seven years of our sample.

5.4 Conclusions

In this chapter we performed an analysis of the relationship between stock return volatility and trading volumes. We showed that a SV model naturally arises when we assume that returns are generated by the Mixture of Distributions Hypothesis (MDH). A natural implication of the MDH is that daily trading volume plays the role of a scale factor for conditional volatility. We derived a SV model that explicitly incorporates daily trading volume in the measurement equation. Our empirical results are based on a sample of five individual stocks and the S&P 500 composite stock index. All models were estimated by Quasi Maximum Likelihood. Results show that including volume leads to a better fit of the models, implying that trading volume is a significant scale factor for the conditional volatility: the role we hypothesized for trading volume is empirically confirmed. We do not find, however, that the persistence of conditional volatility diminishes once volume has been added. This contrasts findings in Lamoureux and Lastrapes (1990), but is supported by results in Bauer and Nieuwland (1994). It may be that trading volume interacts with stock returns and volatility in another way. No conclusive evidence was found for the leverage effect, although some simple tests show that it is predominantly a tail phenomenon. Only for the S&P 500 index we find a significant leverage effect that is much less dominated by outliers.

Appendix 5: VAR Analysis

A useful way to describe the joint dynamics in a system of time series is to estimate Vector AutoRegressive (VAR) models. These models were first introduced by Sims (1980). A VAR(p) model can be written as:

$$x_t = \gamma + \sum_{i=1}^p A_i x_{t-i} + u_t, \quad t = 1, \dots, T.$$

The vector $x_t = (x_{1t}, \dots, x_{Kt})'$ is a $(K \times 1)$ vector that stacks K time series. The $(K \times 1)$ vector γ is a vector with intercepts and the $(K \times K)$ matrices A_i ($i = 1, \dots, p$) contain further coefficients. The K -dimensional innovation process $u_t = (u_{1t}, \dots, u_{Kt})'$ has the properties $E(u_t) = 0$, $E(u_t u_t') = \Sigma$, and $E(u_t u_s') = 0$ for $t \neq s$. Estimation of VAR processes can be done by performing Least Squares on each of the K time series (Lütkepohl, 1991).

The data series that we use in the estimation procedure are the percentage return (y_t), the daily logarithmic trading volume ($\ln V_t$), and a proxy variable for volatility ($\ln y_t^2$).⁷ In a VAR analysis we can investigate the (linear) dynamic interactions between these variables. Note that our results can be indicative only due to the exact non-linear relationship between returns and volatility. Our focus will be on the relation that trading volume has on return and volatility. Interesting questions that can be raised in this context are related to, among others, the lead/lag relations of volume, and the impact of shocks in volume on return and volatility. As we show in the text of this chapter, there is evidence for heteroskedasticity in the volatility series. This could lead to inconsistent estimation of the variance-covariance matrix of the parameters (Hamilton, 1994). This matrix is important for the analysis in the rest of this appendix. We performed a Newey-West procedure in order to correct the covariance matrix of the parameters. We experimented with the lag length, but results did not change substantially. The results below are based on a lag length of 30 in the Newey-West procedure.

The first step in every VAR analysis is the determination of the order p . Lütkepohl (1991) discusses several ways how this can be done. We used two methods: a likelihood-based criterion and the Akaike Information Criterion (AIC). The likelihood method compares the log-likelihood values in a sequential testing procedure. Suppose we know the maximum order M of the VAR. Subsequently, estimate the VAR for $p = M, \dots, 1$ and compute the test statistic

$$\lambda(p) = T \left(\ln |\tilde{\Sigma}_H(M - p)| - \ln |\tilde{\Sigma}_H(M - p + 1)| \right),$$

⁷The volatility series is motivated by the log-square transformation of returns that is used for estimating our stochastic volatility models. LeBaron (1993) uses the same series.

where $\hat{\Sigma}_H(p)$ is the heteroskedastic-consistent maximum likelihood estimator of Σ for the VAR(p) model. The null hypothesis for this test is that the matrix A_p is zero conditional on A_{p+1}, \dots, A_M being zero. The test statistic has an asymptotic $\chi^2(K^2)$ distribution. For our series we set $M = 20$ and we test at a 1% significance level.⁸ The AIC method for determining the order of a VAR model is computed as follows (Lütkepohl, 1991, p.129)

$$AIC(p) = \ln |\hat{\Sigma}_H(p)| + \frac{2pK^2}{T}.$$

The optimal order p is found by minimizing $AIC(p)$. We present the optimal orders for the trivariate VARs in Table 5.9. Remarkably, all orders are around 10, which corresponds with 10 trading days. Maybe this is influenced by a remaining seasonal (*i.e.* every 2 weeks) effect in the series. Our findings are consistent with results in LeBaron (1993) and Hiemstra and Jones (1994), who also perform VAR analyses on systems that include trading volume. In the analyses below we chose the largest order for each VAR system.

TABLE 5.9: Optimal VAR Orders

| | ALC | AMO | BRI | CHA | HEW | S&P 500 |
|--|-----|-----|-----|-----|-----|---------|
| Trivariate VAR: $x_t = (y_t, \ln y_t^2, \ln V_t)'$ | | | | | | |
| Likelihood | 10 | 13 | 11 | 11 | 11 | 12 |
| AIC | 9 | 12 | 10 | 10 | 10 | 11 |

Notes: All entries are orders p for the VAR(p) model. *Likelihood* refers to the likelihood based test procedure described in the text. The significance level is 1%. *AIC* is the Aikake Information Criterion described in the text.

The optimal orders found for all VAR models prohibit the presentation of the estimated coefficients. Furthermore, in higher order ($p > 1$) VARs the dynamic interactions among the variables cannot be readily inferred by looking at the parameters. A more useful technique for analyzing the dynamics is to compute impulse response functions (IRFs). An IRF describes the reaction of one variable in x_t to a shock in any of the variables in x_t . It can be computed by rewriting the VAR as a vector MA(∞) model. Subsequently, we perform a Choleski decomposition on the covariance matrix: $\hat{\Sigma}_H = LL'$, with L a lower triangular matrix. The IRFs are obtained by recursively tracing the effect of L on the coefficients of the MA representation (see Lütkepohl (1991) for an exact exposition of the algorithm). The matrix L summarizes the shocks that we apply to the VAR model. The fact that L is lower triangular means that the ordering of the variables in x_t is of importance, since changing

⁸We repeated the procedure for a 5% significance level, but the results were identical to the ones reported.

the order will result in a different L . A shock in the first variable has contemporaneous effects on the other variables through the elements in the first column of L . A shock in the second variable does not have a contemporaneous effect on the first variable.

We present IRFs for the trivariate VAR systems in figures 5.9 to 5.14. For every stock we have 6 panels. Every panel corresponds with a certain ordering of variables which is described in Table 5.10. Looking at the plots we see that for all series there are substantial cumulative effects of shocks. For Alcoa (Figure 5.9), Amoco (Figure 5.10) and Hewlett-Packard (Figure 5.13) we see that a shock in volume leads to a higher cumulative level of volatility, whatever the ordering of the variables is. For the S&P 500 index series (Figure 5.10) a similar effect is present, but a shock in volume has a negative impact on volatility. We see a persistent effect on returns from a shock in volume for Champion (Figure 5.12) and a negative effect from volume to volatility for all panels.

TABLE 5.10: Description Impulse Response Functions Trivariate VAR

| panel | ordering | Line type | | | |
|-------|----------|-----------|-----------|-------|------------|
| | | solid | long dash | dots | short dash |
| A | (1,2,3) | 1 → 2 | 1 → 3 | 2 → 1 | 3 → 1 |
| B | (2,1,3) | 1 → 2 | 2 → 1 | 1 → 3 | 3 → 1 |
| C | (1,3,2) | 3 → 1 | 1 → 3 | 2 → 1 | 1 → 2 |
| D | (2,3,1) | 1 → 2 | 2 → 1 | 1 → 3 | 3 → 1 |
| E | (3,1,2) | 1 → 3 | 3 → 1 | 1 → 2 | 2 → 1 |
| F | (3,2,1) | 1 → 3 | 3 → 1 | 1 → 2 | 2 → 1 |

Notes: 1 = volume ($\ln V_t$); 2 = volatility ($\ln y_t^2$); 3 = return (y_t). The notation ' $x \rightarrow y$ ' corresponds with the IRF of the variable y resulting from a shock in x .

Another way to investigate the dynamic properties of a VAR system is to test for Granger-causality. A variable y is Granger-causal for another variable z if predictions for z can be made more efficiently by including y in the information set that is used for predicting z . In a VAR system a test for Granger-causality amounts to a joint test for zero coefficients in the A_i ($i = 1, \dots, p$) matrices. For example, if we want to test if the second variable in a VAR Granger-causes the first variable, the null hypothesis is that the elements (1,2) in every A_i are equal to zero. The test statistic turns out to be a standard F-test with degrees of freedom equal to p and $T - pK - 1$. Again we took account of the possible heteroskedasticity in the data when we compute the test statistics.⁹ In Table 5.11 we

⁹The test statistic is given by $F = (1/p)(C\hat{\beta} - c)'[C((ZZ')^{-1} \otimes \hat{\Sigma}_H)C']^{-1}(C\hat{\beta} - c)$. The null hypothesis is that $C\beta = c$. The matrix $Z \equiv (Z_0, \dots, Z_{T-1})$ is the $((Kp+1) \times 1)$ matrix with elements $Z_t \equiv (1, x_t, \dots, x_{t-p+1})$.

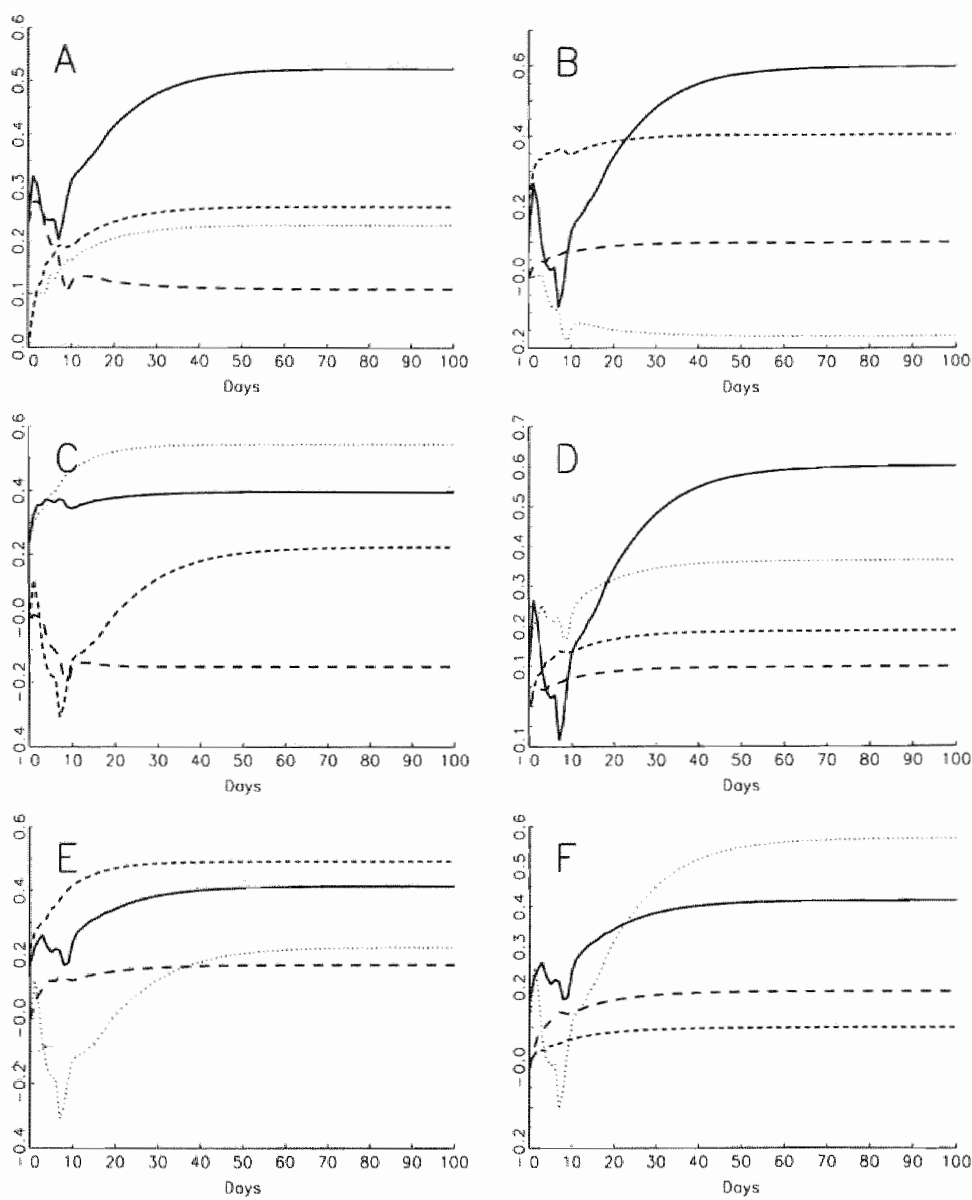


FIGURE 5.9: Impulse Responses Trivariate VAR: Alcoa
See text and Table 5.10 for description.

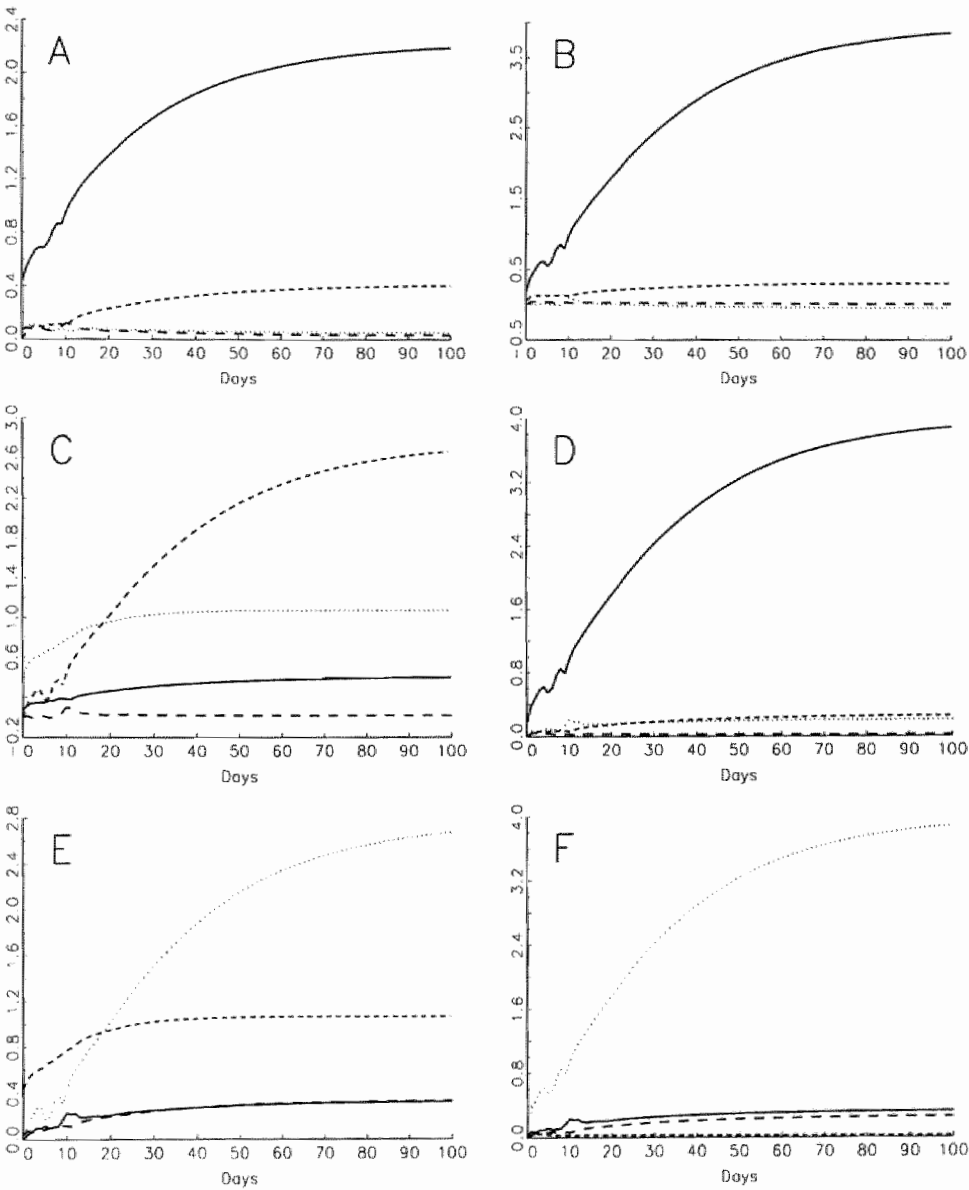


FIGURE 5.10: Impulse Responses Trivariate VAR: Amoco
See text and Table 5.10 for description.

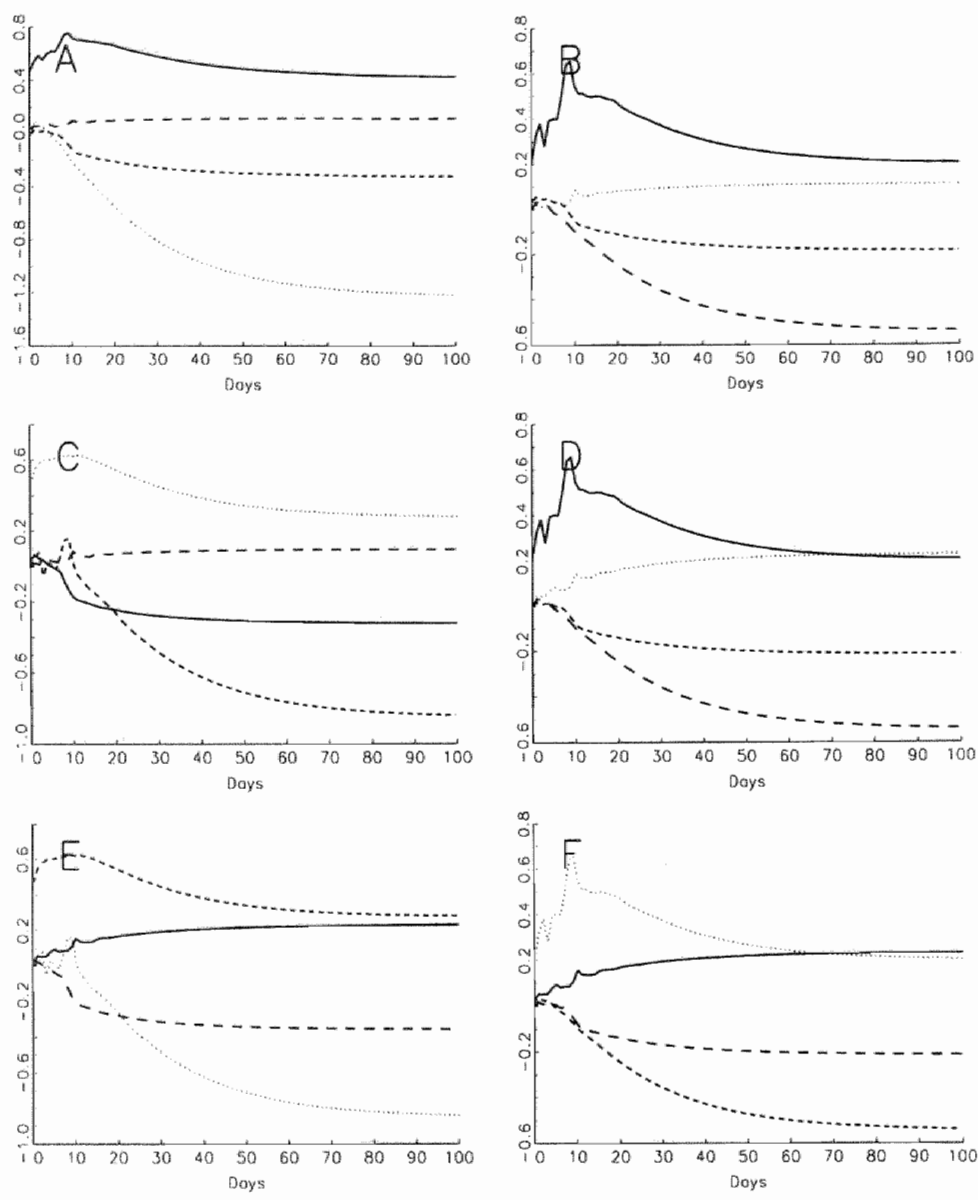


FIGURE 5.11: Impulse Responses Trivariate VAR: Bristol-Myers
See text and Table 5.10 for description.

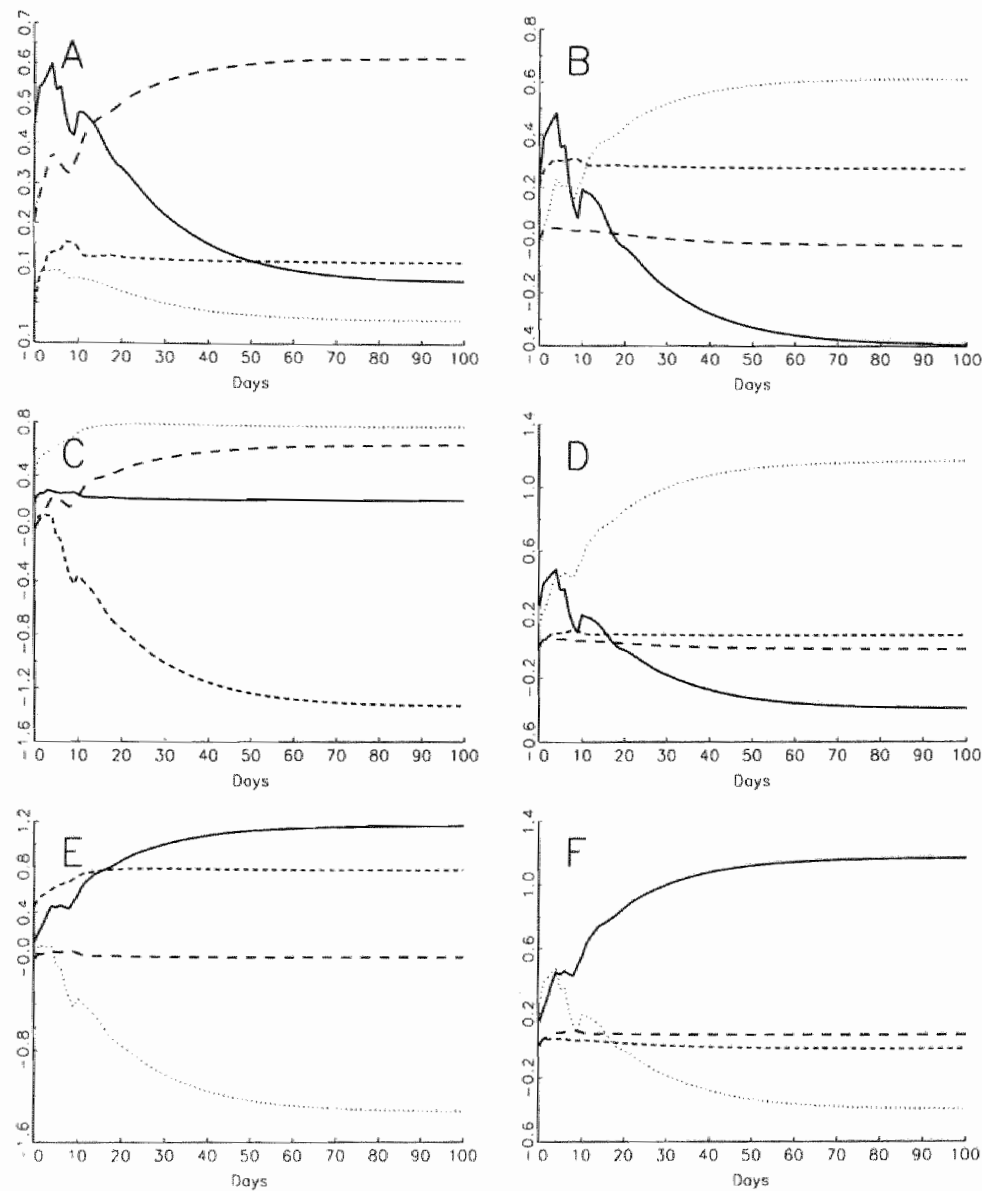


FIGURE 5.12: Impulse Responses Trivariate VAR: Champion
See text and Table 5.10 for description.

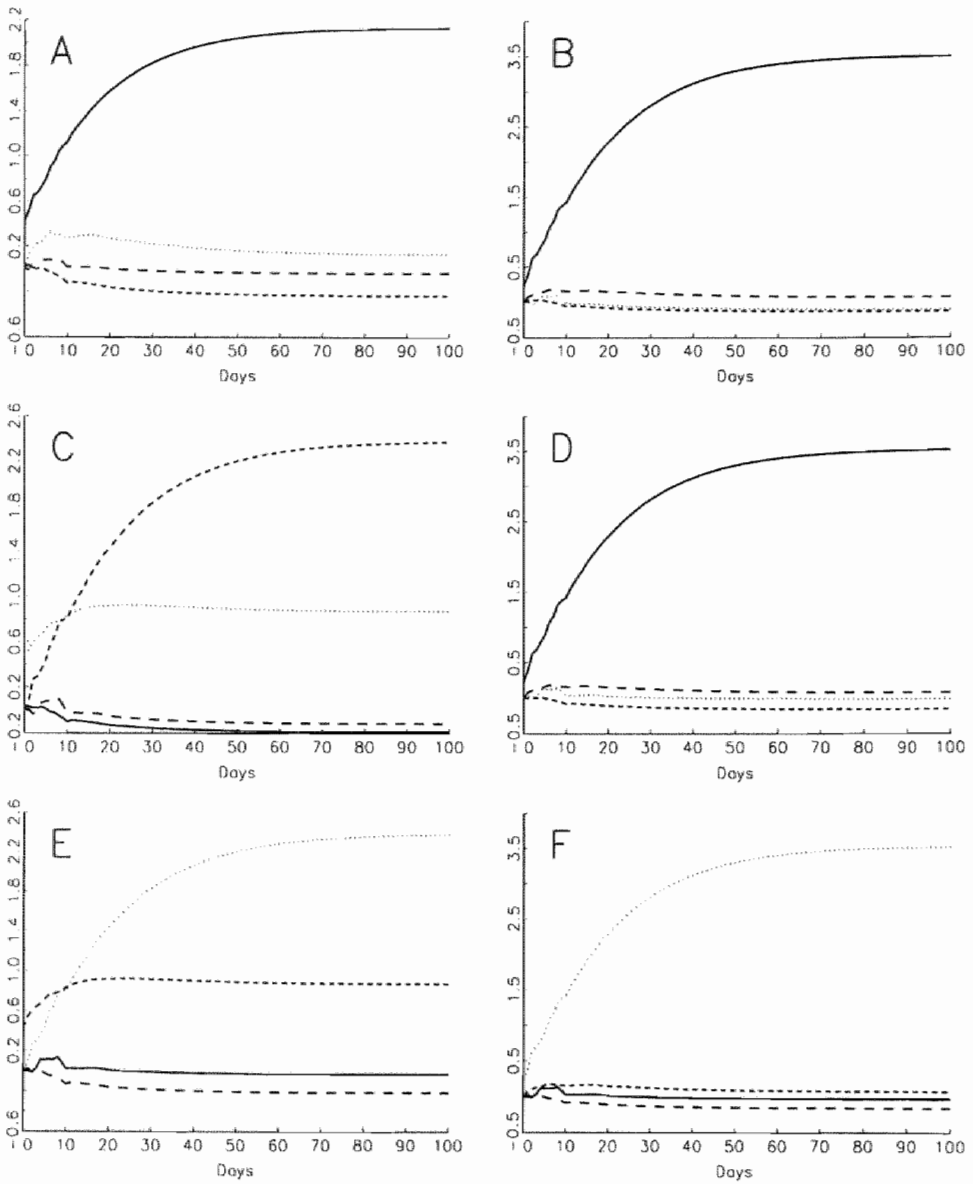


FIGURE 5.13: Impulse Responses Trivariate VAR: Hewlett-Packard
See text and Table 5.10 for description.

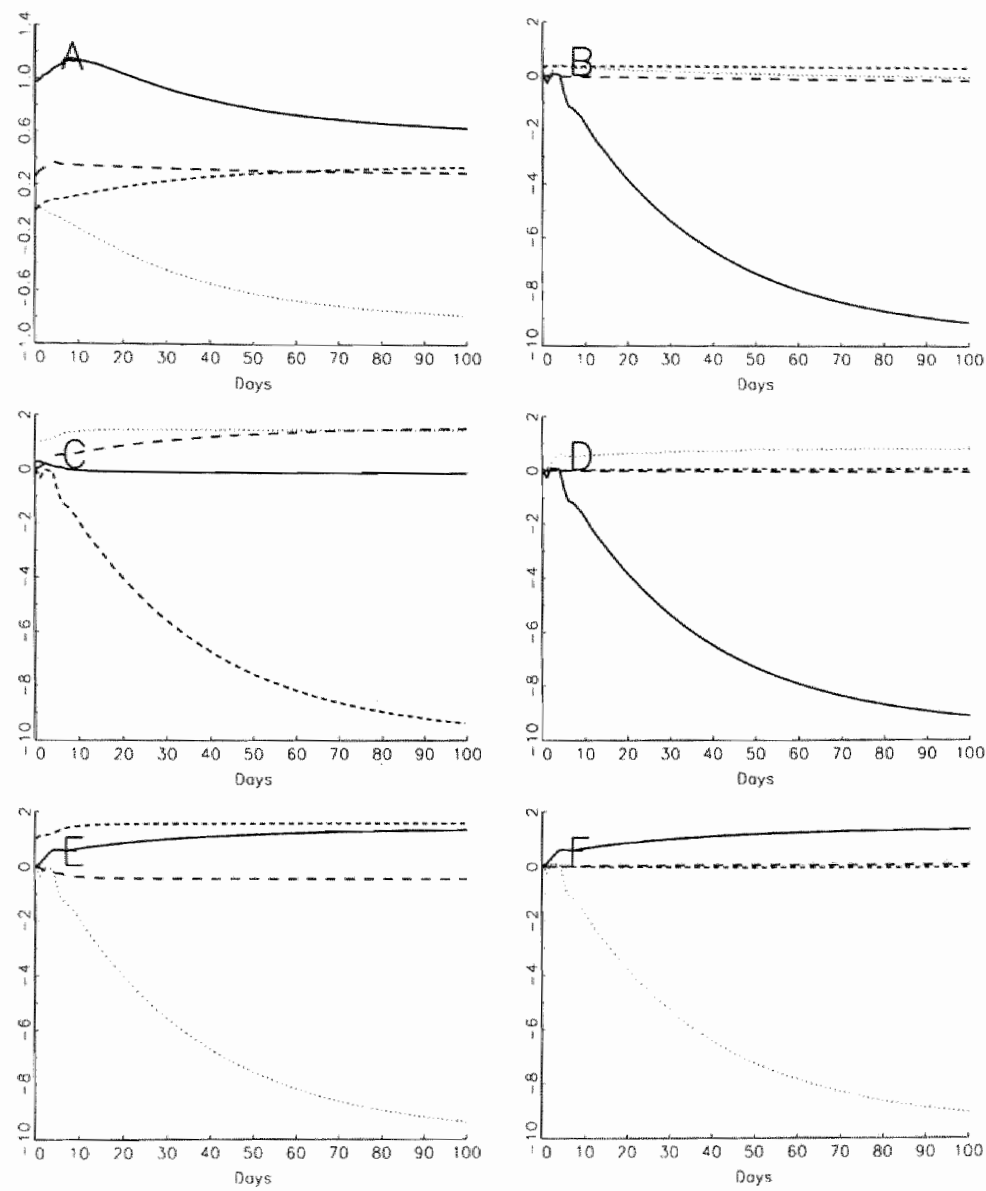


FIGURE 5.14: Impulse Responses Trivariate VAR: S&P 500
See text and Table 5.10 for description.

report results from doing Granger-causality tests on the trivariate VARs.

TABLE 5.11: Granger-Causality Tests on Trivariate VAR

| Alcoa | | | | Amoco | | | |
|-------|---|---|---|-------|---|---|---|
| from | | | | from | | | |
| | 1 | 2 | 3 | | 1 | 2 | 3 |
| to 1 | • | • | • | to 1 | • | • | • |
| to 2 | • | • | • | to 2 | • | • | • |
| to 3 | • | • | • | to 3 | • | • | • |

| Bristol-Myers | | | | Champion | | | |
|---------------|---|---|---|----------|---|---|---|
| from | | | | from | | | |
| | 1 | 2 | 3 | | 1 | 2 | 3 |
| to 1 | • | • | • | to 1 | • | • | • |
| to 2 | • | • | • | to 2 | • | • | • |
| to 3 | • | • | • | to 3 | • | • | • |

| Hewlett-Packard | | | | S&P 500 | | | |
|-----------------|---|---|---|---------|---|---|---|
| from | | | | from | | | |
| | 1 | 2 | 3 | | 1 | 2 | 3 |
| to 1 | • | • | • | to 1 | • | • | • |
| to 2 | • | • | • | to 2 | • | • | • |
| to 3 | • | • | • | to 3 | • | • | • |

Notes: 1 = volume ($\ln V_t$); 2 = volatility ($\ln y_t^2$); 3 = return (y_t). A • means that the column variable Granger-causes the row variable at a 1% significance level.

From the table it can be seen that all diagonal elements are significant. This means that the lag history of every variable is important in predicting its own current value. More interesting is that in 4 cases it seems to be that volatility causes volume, and in two cases past returns Granger-cause volume. It is never the case that volume Granger-causes either volatility or returns, although contemporaneous correlations show that volume is significantly correlated with the other two variables. This contrast with some of the findings from the impulse response functions, where we found tentative evidence that shocks to volume lead to persistence effects in the two other variables.¹⁰

The $((K^2p+K) \times 1)$ vector β stacks the coefficients in the VAR: $\beta \equiv \text{vec}(\gamma, A_1, \dots, A_p)$. Furthermore, define x as $\text{vec}(x'_1, \dots, x'_T)'$ and $u \equiv \text{vec}(u'_1, \dots, u'_T)'$. With this notation the VAR(p) model can be written as $x = (Z' \otimes I_K)\beta + u$, with I_K an identity matrix.

¹⁰Note that we have not computed any standard errors on the IRFs. So care has to be taken in relating the IRFs to the results from the Granger-causality tests.

In conclusion we can say that sizable dynamic effects related to trading volume seem to be present in the VAR systems we analyzed. The VAR models do not give us any economic rationale about why volume should be incorporated. They only indicate that volume is an important variable in a model that tries to describe the way stock returns and return volatility behave.

Chapter 6

Hedging Foreign Currency Exposure: Do Fat Tails Matter?

In the previous chapters our focus has been on discussing models that can describe the volatility of financial assets. For financial decision-making volatility models provide an important ingredient for issues concerning portfolio allocation, hedging, and the pricing of derivative assets. In this chapter as well as in the next chapter we will look at two applications in international finance in which volatility plays an important role. In this chapter we will turn our attention to hedging foreign currency exposure in portfolios that consist of both a domestic investment as well as a foreign investment.

A long debated issue in international finance is whether or not foreign currency exposure should be hedged. On the one hand several studies have shown that hedging does improve portfolio performance in the sense that it improves the risk-return trade-off of the portfolio.² These results build on the fact that exchange rate changes in excess of the forward discount average about zero and have low correlations with either local and foreign asset returns. This means that the currency part of an international portfolio consists mainly of independent risk that can readily be hedged away at low return sacrifices (Froot, 1993).

On the other hand people claim that currency exposure should not be hedged at all. For example, Froot (1993) shows that hedging does not lead to lower portfolio risks over long horizons. In some circumstances it might happen that a fully hedged portfolio leads to a larger return variance than an unhedged equivalent.

International asset pricing and equilibrium currency hedging have been extensively ana-

²See, for example, Glen and Jorion (1993) and Levich and Thomas (1993) for recent evidence.

lyzed in a theoretical sense. Seminal articles include among others Solnik (1974), Sercu (1980), Anderson and Danthine (1981), Stulz (1981) and Adler and Dumas (1983). The international asset pricing models derived in these articles indicate that in equilibrium all investors should hold a combination of their national risk-free asset in combination with the world market portfolio partly hedged against currency exposure. More recently, Black (1990) introduced the method of so-called universal currency hedging. Based on the assumptions that every investor has the same risk tolerance and that national wealth is equal to the value of national assets, *i.e.* on balance there is no foreign investment, Black argues that every investor should hedge the same percentage, irrespective of the home country and moreover that this hedge should always be less than one. Black's analysis on universal hedging has led to considerable discussion on the applicability of universal hedge ratios in the literature (Solnik, 1993, and Adler and Prasad, 1992).

Until recently, empirical applications of currency hedging were often based on the so-called regression hedge or minimum-variance hedge.³ The optimal hedge ratio under minimum variance is equal to the ratio of the covariance between the unhedged return on the portfolio and the return on a futures position on the one hand and the variance of the futures position on the other hand. The optimal hedge ratio is thus equal to the coefficient of a regression of the unhedged portfolio return on the return of the futures position.

However, both the theoretical and empirical models of currency hedging do not take into account one of the most salient stylized facts of financial markets that have been reported in the literature: the excess kurtosis in unconditional asset and exchange rate returns distributions. In general, unconditional distributions of financial time series have fatter tails and higher peaks than the normal distribution.⁴ Our aim in this chapter is to describe optimal hedge strategies under fat-tailed distributions of the relevant returns.

The outline of this chapter is as follows. In Section 6.1 we discuss the case for currency hedging in international investments. We will also present some descriptive statistics on the dataset we use and introduce some notation. In Section 6.2 the implications of optimal hedging under fat-tailed distributions is investigated. Section 6.3 contains some concluding remarks.

³See Anderson and Danthine (1981) Eaker and Grant (1987) and Solnik (1988)

⁴See Boothe and Glassman (1987), Diebold (1988), Koedijk, Schafgans and De Vries (1990) and De Vries (1992).

6.1 Currency Hedging Revisited

International portfolio diversification has gained an ever growing interest from financial researchers. An important reason for this is that by "going abroad" investors can obtain more favorable characteristics of their portfolios (Levy and Sarnat, 1970). In practice, however, many large investors still hold a portfolio that is heavily biased towards home assets. It has been hypothesized that transactions' cost, regulatory restrictions, non-traded goods and investor irrationality are candidates for explaining the home-bias (Glen and Jorion, 1993).⁵ The recent turbulence (and corresponding increased volatility) in world currency markets may have deterred investors from investing abroad. However, when an international portfolio has been well-hedged against currency exposure, increased exchange rate volatility should have no major influence on whether portfolio weights have to be adjusted *ceteris paribus*.

When choosing an internationally diversified portfolio an investor should consider which fraction of the foreign currency exposure he wants to hedge. This can be done in several ways. The most investigated (and applied) hedging strategy is the so-called currency overlay management strategy. The optimal hedges are determined conditional on the portfolio allocation. Clearly this strategy is suboptimal compared to a more general hedging strategy that incorporates the hedging decisions simultaneously with the portfolio allocation (Glen and Jorion, 1993). Nevertheless, the currency overlay management strategy has been applied widely in practice. In this chapter we will consider the hedging decisions conditional on some portfolio allocation. For simplicity we will focus on bivariate hedges, *i.e.* we consider portfolios that consist of a domestic investment and one foreign investment only.

The consequences of exchange rate volatility on a foreign asset may be studied in the following way. In this chapter all computations will be done from the perspective of a U.S. investor. In other words, the U.S. is the home country. Let $S_{i,t}$ be the spot dollar price of foreign currency i at time t , and $P_{i,t}$ be the nominal local currency value of a portfolio in country i including dividend payments. Suppose we are at time t . The nominal exchange

⁵Another explanation for the home-bias might be given by the international portfolio model of Adler and Dumas (1983). Investors from different countries consume different bundles of goods. Inflation risk and violations of Purchasing Power Parity (PPP) induce investors from different countries to hold different portfolios. Consequently, Adler and Dumas (1983) argue that the home-bias might be caused by the fact that domestic investments form a hedge against inflation. Cooper and Kaplanis (1994) build on these observations by making an international portfolio model that explicitly models PPP deviations and some other costs. They find that hedging inflation risk alone cannot account for the home bias.

rate return between times t and $t + 1$ is defined as

$$s_{i,t+1} \equiv \frac{S_{i,t+1}}{S_{i,t}} - 1 \approx \ln S_{i,t+1} - \ln S_{i,t}.$$

Note that the exchange rate one period ahead is a random variable, whereas the current exchange rate is known to the investor. The logarithmic approximation of the return is used in the empirical computations later on. Equivalently, define the local return on an investment as

$$r_{i,t+1} \equiv \frac{P_{i,t+1}}{P_{i,t}} - 1 \approx \ln P_{i,t+1} - \ln P_{i,t}.$$

This return could be both a dollar return on a U.S. investment or a local currency return on a foreign investment. Consequently, we can write the dollar return on the foreign investment as

$$r_{i\$,t+1} = (1 + r_{i,t+1})(1 + s_{i,t+1}) - 1.$$

Again, this return is computed at time t and is a random variable. In many practical studies cross products of returns are often skipped, because they are of a lower order of magnitude. As long as returns are relatively small, the error made is very small. To simplify the exposition, we will adhere to this practice in the empirical parts of this chapter. The last equation may thus be approximated as

$$r_{i\$,t+1} \approx r_{i,t+1} + s_{i,t+1}.$$

This equation shows that the exchange rate return is an integral part of the dollar return on an international investment. Likewise, the contribution that exchange rates have to the variance of the whole portfolio can be determined. The variance of the dollar return of an investment in country i is given by

$$\text{var}(r_{i\$,t+1}) \approx \text{var}(r_{i,t+1}) + \text{var}(s_{i,t+1}) + 2\text{cov}(s_{i,t+1}, r_{i,t+1}).$$

Both the variances of local asset returns and exchange rate returns, as well as the covariance between these two variables are important.

The portfolios that are under investigation in this chapter consist of a domestic and one foreign component. Let $w_{i,t}$ be the fraction of the portfolio that is invested in country i , and let $w_{\$,t}$ be the proportion of wealth that is invested at home. These weights sum to one: $w_{i,t} + w_{\$,t} = 1$. Let $r_{p\$,t+1}^i$ be the random return in dollars on the portfolio $w_{i,t}r_{i\$,t+1} + w_{\$,t}r_{\$,t+1}$, where $r_{\$,t+1}$ is the dollar return on a domestic (U.S.) investment. Then

$$\begin{aligned} \text{var}(r_{p\$,t+1}^i) &= \sum_{k=i,\$} \sum_{j=i,\$} w_{k,t} w_{j,t} \text{cov}(r_{k,t+1}, r_{j,t+1}) + \\ &\quad w_{i,t}^2 \text{var}(s_{i,t+1}) + \\ &\quad 2 \sum_{k=i,\$} w_{k,t} w_{i,t} \text{cov}(r_{k,t+1}, s_{i,t+1}). \end{aligned} \tag{6.1}$$

In order to get a feeling for what these numbers are in a real-life setting we computed each of the covariances in (6.1) using weekly data from DATASTREAM. We use country stock indices and exchange rates quoted at Wednesdays. The countries that we included in our sample are the United States, the United Kingdom, Japan and Germany. The stock index series are total return series computed by DATASTREAM. The exchange rates are quoted in Pound Sterling and were converted to U.S. dollars. The sample period starts at January 2, 1980 and ends on May 31, 1995, giving us 805 observations. Note that during this whole period exchange rates were fluctuating freely, except for the two European countries whose currency rates were tied together in the European Exchange Rate Mechanism (ERM) from 1979 to 1992. We present some summary statistics of both stock index returns and exchange rate returns in Table 6.1.

The minimum value for every stock index is obtained in October 1987. The stock returns show a considerable amount of negative skewness and excess kurtosis, which is consistent with the stylized facts of stock market returns. Exchange rate returns show much less evidence for skewness and kurtosis. There is not much evidence for serial correlation in any of the returns or exchange rate series. However, for all the stock return series the Ljung-Box statistics for serial correlation in the squares of the series are significant at the 5% significance level, although the first 5 autocorrelations differ significantly from zero in only a few cases. The Augmented Dickey-Fuller (ADF) test statistics show that there does not appear to be any evidence for unit roots in any of the series.

In Table 6.2 we present simple correlation coefficients for asset and exchange rate returns. Two main conclusions can be drawn from the table. First, correlations within an asset class are positive and high.⁶ The ongoing integration of international capital markets might be responsible for this result. In case of foreign exchange rates an additional explanation is the numeraire effect (see also Chapter 4 of this thesis). Every exchange rate in the tables 6.1 and 6.2 is quoted in U.S. dollars, leading to equal responses of the rates to U.S. news events. Note the high correlation of the two ERM currencies. A second observation that can be made is that the cross-correlations between stock markets on the one hand and foreign exchange rate markets on the other hand, is remarkably low. Compared to Eun and Resnick (1988) who use weekly data from 1980 to 1985, we find qualitatively the same results, with the exception that our stock return data have higher correlations and our cross-correlations are lower. Another difference is that we sometimes find negative cross-correlations between exchange rate and stock returns implying that exchange rate movements offset local stock market movements.

A comparison of asset and exchange rate returns in Table 6.1 shows that the standard

⁶The standard error is $1/\sqrt{804} \approx 0.035$ (see Granger and Newbold, 1986).

TABLE 6.1: Descriptive Statistics Stock Index Returns and Exchange Rate Returns

| | Stock index returns $r_{i,t+1}$ | | | | Exchange rate returns $s_{i,t+1}$ | | | |
|-----------|---------------------------------|---------|---------|---------|-----------------------------------|--------|--------|--|
| | US | UK | JP | GE | UK | JP | GE | |
| mean | 0.277 | 0.346 | 0.152 | 0.205 | -0.042 | 0.129 | 0.024 | |
| st. dev. | 2.059 | 2.148 | 2.399 | 2.189 | 1.625 | 1.471 | 1.627 | |
| min. | -16.26 | -17.34 | -10.45 | -15.97 | -8.669 | -4.999 | -7.314 | |
| max. | 8.059 | 10.12 | 13.62 | 6.986 | 7.496 | 6.518 | 7.936 | |
| skewness | -0.860 | -1.173 | -0.351 | -1.091 | -0.173 | 0.541 | 0.156 | |
| kurtosis | 9.301 | 13.45 | 6.443 | 8.740 | 5.793 | 4.424 | 4.487 | |
| ADF(4) | -13.38* | -12.12* | -11.86* | -12.27* | -11.36 | -11.47 | -11.81 | |
| $\rho(1)$ | 0.028 | 0.136* | 0.029 | 0.125 | 0.028 | 0.059 | 0.030 | |
| $\rho(2)$ | -0.024 | 0.063* | 0.076* | 0.115 | -0.003 | 0.105* | 0.041 | |
| $\rho(3)$ | 0.038 | 0.009 | 0.071* | 0.069 | -0.004 | 0.034 | -0.027 | |
| $\rho(4)$ | -0.047 | -0.001 | -0.018 | -0.016 | -0.055 | 0.026 | 0.055 | |
| $\rho(5)$ | -0.041 | -0.003 | 0.006 | -0.036 | 0.061 | 0.018 | 0.014 | |
| LB(20) | 30.42* | 34.29* | 32.07* | 39.48* | 25.64 | 31.37 | 16.83 | |

Notes: All return series are percentages per week. Foreign exchange returns are computed as U.S. dollar returns: Weekly data; sample period January 2, 1980 – May 31, 1995. $ADF(4)$ is the Augmented Dickey-Fuller test for a unit root with 4 lags. The McKinnon 5% critical level is -2.866. $\rho(n)$ is the n th-order autocorrelation. Critical values are given by $\pm 2/\sqrt{804} \approx 0.071$. $LB(20)$ is the Ljung-Box test with 20 degrees of freedom. Its critical value is 31.4 at a 5% significance level. A * means significance at 5% level.

TABLE 6.2: Correlation Matrices

| | UK | JP | GE | US |
|--|--------|--------|--------|-------|
| Stock Market Returns in Local Currencies | | | | |
| U.K. | | 0.306 | 0.398 | 0.522 |
| Japan | | | 0.276 | 0.352 |
| Germany | | | | 0.395 |
| Exchange Rate Changes against the U.S. Dollar | | | | |
| U.K. | | 0.502 | 0.771 | — |
| Japan | | | 0.620 | — |
| Stock/Exchange Cross-correlations | | | | |
| U.K. | -0.044 | 0.000 | -0.079 | — |
| Japan | 0.072 | 0.118 | 0.031 | — |
| Germany | -0.118 | -0.058 | -0.121 | — |
| U.S. | 0.025 | 0.014 | 0.034 | — |

Notes: Weekly data; sample period January 2, 1980 – May 31, 1995. Each row in the Stock/Exchange Cross-correlations part shows correlations of the stock return with each of the dollar denominated exchange rate returns.

deviations of U.S. dollar exchange rates are only slightly lower than the variances of local stock returns. A useful way to get a feeling for the influence of currency fluctuations on the total portfolio variance is to compute the contributions of each of the components in Equation (6.1). Assume that the investor invests his money equally in the home and the foreign country, *i.e.* $w_{i,t} = w_{\$,t} = 0.5$, for all i and t . In Table 6.3 we have listed the decomposition results. The upper two lines show the mean returns for the local stock indices and exchange rates separately. Currencies do not add much return for the equally weighted portfolios over this sample period. The lower three lines show the decomposition of $\text{var}(r_{p,\$,t+1}^i)$. Currencies are a major part of the variance of a U.S. dollar denominated portfolio. The negative cross-correlation means that exchange rate movements offset local return fluctuations.⁷

A useful strategy to remove exchange risk from the portfolio is to sell the expected currency proceeds forward. The unexpected part of the dollar proceeds remains unhedged. Let $F_{i,t}$ be the 1-week forward exchange rate. When an investor sells U.S. dollars forward against another currency, he is able to buy dollars back at a fixed rate one week from now. If he

⁷We also computed the portfolio decompositions for several subperiods. The conclusions are qualitatively the same.

TABLE 6.3: Portfolio Decompositions for 3 Bilateral Portfolios

| Component | Contributions | | | |
|---|----------------|---------------|----------------|--|
| | UK | JP | GE | |
| $\sum_{k=i,\$} \frac{1}{2} \mu(r_k)$ | 0.311 (107.4%) | 0.214 (76.9%) | 0.241 (95.3%) | |
| $\frac{1}{2} \mu(s_i)$ | -0.021 (-7.4%) | 0.064 (23.1%) | 0.012 (23.1%) | |
| $\sum_{k=i,\$} \sum_{j=i,\$} \left(\frac{1}{2}\right)^2 \text{cov}(r_k, r_j)$ | 3.368 (84.3%) | 3.368 (81.4%) | 3.148 (86.2%) | |
| $\frac{1}{4} \text{var}(s_i)$ | 0.660 (16.5%) | 0.541 (13.1%) | 0.662 (18.1%) | |
| $2 \sum_{k=i,\$} \left(\frac{1}{2}\right)^2 \text{cov}(r_k, s_i)$ | -0.035 (-0.8%) | 0.231 (5.5%) | -0.160 (-4.3%) | |

Notes: Weekly data; sample period january 2, 1980 – May 31, 1995. Returns are measured in percentages per week. Numbers between parentheses are percentage contributions with respect to the total portfolio mean and variance. $\mu(\cdot)$ is the sample mean. All moments are unconditional. The index i corresponds with the foreign countries ($i=UK, JP, GE$).

simultaneously buys foreign currency on the spot market, the initial investment is zero and the investor incurs no risk at time t . This strategy of selling a currency forward and at the same time purchasing a foreign currency spot is called a currency swap. Note that the proceeds of a currency swap can be replicated by taking offsetting positions in dollar and foreign currency deposits.⁸ This statement is also called covered interest rate parity

$$f_{i,t} + 1 \equiv \frac{F_{i,t} - S_{i,t}}{S_{i,t}} + 1 = \frac{I_{\$,t} + 1}{I_{i,t} + 1},$$

where $I_{\$,t}$ and $I_{i,t}$ are the interest rates in percentage points on 1-week deposits denominated in dollars and a foreign currency, respectively. The term $f_{i,t}$ is also called the forward exchange premium. In logarithmic terms covered interest rate parity can be written as

$$f_{i,t} \approx \ln F_{i,t} - \ln S_{i,t} = I_{\$,t} - I_{i,t} \quad (6.2)$$

We assume that the deposit interest rates are known at time t implying that the basis is fixed for the forward contract.⁹ Note that covered interest rate parity is an arbitrage condition. In other words, if $F_{i,t} > S_{i,t}$ then $I_{\$,t} > I_{i,t}$ meaning that an expected dollar depreciation is offset by a higher U.S. interest rate to ensure equilibrium.

⁸More specifically, the investor should take a short position in the foreign currency and an offsetting long position in the U.S. dollar.

⁹This assumption is harder to maintain for futures contracts, for which daily settlements occur. See Solnik (1988).

Hedging foreign currency exposure through forward contracts can be performed by transforming the expected dollar proceeds $E(r_{i,t+1} + s_{i,t+1})$ to the certain dollar return $Er_{i,t+1} + f_{i,t}$. If the forward premium is positive ($f_{i,t} > 0$) then it is expected that the dollar will depreciate enhancing the total portfolio return in dollars. The unexpected foreign asset return $r_{i,t+1} - Er_{i,t+1}$ is still exposed to a currency movement between time t and $t+1$. The dollar value of the unexpected asset returns is approximately equal to $(r_{i,t+1} - Er_{i,t+1})(1 + s_{i,t+1}) - 1 \approx r_{i,t+1} - Er_{i,t+1}$. This shows that shocks in local returns are approximately unaffected by currency movements. In other words, the *ex ante* quantity risk of the foreign investment is of a lower order of magnitude (Eun and Resnick, 1988).

In order to investigate the effects of currency hedging on a foreign investment we consider the case of a U.S. investor who wishes to hedge his foreign currency exposure in a single foreign currency. His portfolio consists of both foreign and domestic investments in the local equity index, and a number of forward contracts on the exchange rate. We define the amount of foreign currency i sold forward at time t with delivery at time $t+1$ to be $h_{i,t}w_{i,t}/S_{i,t}$. Since forward contracts do not require an initial investment a rate of return on forwards cannot be defined. At time $t+1$ forward contracts give a random payoff of $S_{i,t+1} - F_{i,t}$. We normalize this payoff by dividing by the current exchange rate $S_{i,t}$. Using 6.2 we can approximate the random payoff on a forward contract as

$$\frac{S_{i,t+1} - F_{i,t}}{S_{i,t}} \approx s_{i,t+1} - f_{i,t}$$

In other words, the normalized payoff on a forward contract consists of both a random exchange rate return and the interest rate differential.

With the definitions above the U.S. dollar return on an international portfolio consisting of a U.S. investment and an investment in country i with forward contracts becomes

$$\begin{aligned} r_{t+1}^h &\equiv w_{i,t}r_{i,t+1} + w_{\$,t}r_{\$,t+1} - h_{i,t}w_{i,t}(s_{i,t+1} - f_{i,t}) \\ &= w_{i,t}r_{i,t+1} + w_{\$,t}r_{\$,t+1} + w_{i,t}(1 - h_{i,t})s_{i,t+1} + h_{i,t}w_{i,t}f_{i,t} \end{aligned} \quad (6.3)$$

$$\equiv wr_{t+1} + (1 - w)r_{t+1}^* + (1 - w)(1 - h)s_{t+1} + (1 - w)hf_t. \quad (6.4)$$

To ease the exposition in the rest of this chapter we have redefined the relevant variables in the last line of (6.4). The foreign asset return is given by r_{t+1}^* . The portfolio weight w and the hedge ratio h do not depend on time index t anymore, since we will concentrate on unconditional portfolio and hedging strategies only. Since the investor can only invest in two countries we only need one weight w to fix the portfolio allocation. The random return on the portfolio is equal to the return on the local equity investments plus the partly hedged exchange rate return and the cost of hedging. The variable h is called the *hedge ratio*. If we set $h = 1$ the portfolio return does not depend on exchange rate movements anymore.

Alternatively, if $h = 0$ the portfolio returns are fully exposed to exchange rate risk. The cost of hedging the portfolio consists of the interest rate difference. A hedging strategy consists of determining finding optimal hedge ratios h given some objective function. In the rest of this chapter we will turn to this problem.

6.2 Optimal Hedge Ratios

In this section we will perform an analysis of the computation of optimal hedge ratios in several circumstances. The computation of optimal hedge ratios does not differ considerably from the computation of optimal portfolio weights. The portfolio of local returns (domestic and foreign) and the exchange rate is augmented with the payoffs on a forward contract. Consequently, by maximizing an expected utility function with respect to the portfolio weight and the hedge ratio, optimal values for w and h can be computed. The choice of the utility function relates to the moments of the underlying returns distributions, that an investor deems important. For example, the much-used mean-variance utility function in finance implies that only the first two moments of returns are relevant in determining the optimal weights of a portfolio.

As was stated before, one of the most salient characteristics of financial asset returns is the leptokurtosis or fat-tailedness in the unconditional distributions. Two different views have been taken to explain this phenomenon. The first view suggests that the observations were drawn from a distribution that changes over time. This explanation is related to the conditional volatility models that we discussed earlier in this thesis. The choice of a conditional error distribution in an ARCH or SV model is not straightforward, though. In an ARCH model a conditional normal distribution is frequently chosen. However, in this case the standardized residuals still exhibit considerable leptokurtosis in many cases.¹⁰ Other candidates for the error distribution have been given by Bollerslev (1988), who proposed to use a scaled Student t distribution as the conditional error distribution, and Nelson (1991), who used a General Error Distribution.

The second view suggests that the observations are drawn independently from a distribution that has a density with fatter tails than a normal distribution. Useful candidates that have been proposed in the literature are the Student t distribution (Rogalski and Vinso, 1977), stable Paretian distributions (Boothe and Glassman, 1987;), mixtures of normals (Boothe and Glassman), and mixed-jump processes (Akgriray and Booth, 1988; Tucker and Pond, 1988). Results from comparisons among these candidate distributions do not

¹⁰See, for example, Hsieh (1988).

indicate a clear preference for any of these distributions. Sample period and the type of data (foreign exchange, common stock) seem to be important in choosing the best-fitting unconditional distribution.

The calculation of optimal hedge ratios can be performed under both views of the origin of unconditional leptokurtosis. With respect to the first view, Cecchetti, Cumby and Figlewski (1988), Baillie and Myers (1991), and Kroner and Sultan (1993) develop dynamic hedge models based on the GARCH methodology. In these models the hedge ratios are computed by optimizing a mean-variance expected utility objective function using the conditional moments that are implied by a (multivariate) GARCH model. Consequently, hedge ratios change over time. Here we will concentrate on the computation of hedge ratios (and portfolio weights) under unconditional leptokurtic distributions. We will leave a comparison between the two competing views on the origins of leptokurtosis for future research.

A very common utility function in finance is the mean-variance expected utility function:

$$EU(x) = E(x) - \lambda \text{Var}(x)$$

The parameter λ is a risk aversion parameter. The higher this parameter, the more risk averse the investor is. In order to obtain optimal hedge ratios we would have to maximize this function with respect to the hedge ratios and solve the first-order conditions. Anderson and Danthine (1981) determine the hedge strategy conditional on the portfolio allocation given by the weights w . They find that the optimal hedge in a general case amounts to

$$h = \frac{1}{2\lambda} \Sigma_{gg}^{-1} \mu_g + \Sigma_{gg}^{-1} \Sigma_{gr} x, \quad (6.5)$$

where Σ_{gg} is the covariance matrix of the payoffs on the forward contracts, μ_g is the expected payoff on the forwards, x is a portfolio weight vector and Σ_{gr} is the covariance matrix between the forward payoffs and the underlying investments. In the two-country case that we are considering we would get the following equivalents

$$\begin{aligned} \mu_g &= E(s_{t+1} - f_t) \\ \Sigma_{gg} &= \text{Var}(s_{t+1} - f_t) \\ \Sigma_{gr} &= \left(\text{Cov}(r_{t+1}, s_{t+1} - f_t), \text{Cov}(r_{t+1}^*, s_{t+1} - f_t) \right)' \\ x &= (w, 1 - w)' \end{aligned}$$

Note that we have kept the weights w fixed in this optimization. This is a common practise in the hedging literature. The optimal hedge ratio in (6.5) consists of two parts. The first part can be considered to be a speculative part, since it includes the expected payoff on the hedges. The second part is the "pure" hedge component, because it minimizes the variance

of the hedged portfolio return r_{t+1}^h . Aptly, this component is often called the minimum variance hedge ratio. Note that this minimum variance hedge ratio can also be found by doing a regression of the hedge payoff $(1-w)(s_{t+1} - f_t)$ on the unhedged portfolio returns $wr_{t+1} + (1-w)(r_{t+1}^* + s_{t+1})$.

As an illustration we show combinations of sample means and sample variances of r_{t+1}^h as a function of the hedge ratio h in Figure 6.1. The data were described in the previous section. We used the whole sample period from 1980 to 1995. All computations are made from the viewpoint of an American investor. The hedge ratios chosen are in the interval $[0, 1]$ with an increment of 0.1. Every dot corresponds with one particular hedge ratio. We have set the portfolio weights equal to one half. ($w = 0.5$). In all three plots we find that the lowest dot corresponds with a unit hedge ratio and the highest dot with an unhedged position ($h = 0$). Thus, hedging lowers expected return, but at the same time diminishes the variance of the portfolio up to some level. It depends on the risk aversion of an investor to what extent he wants to sacrifice expected returns in order to have a lower variance. From the plots we can also see that the minimum variance hedge ratios are close to one for all the countries.

The sample statistics in Table 6.1 showed that there is a considerable amount of leptokurtosis in the data. Intuitively, leptokurtosis is a characteristic of portfolio returns that investors do not like. It means that the probability of a large shock in the return of the portfolio is larger than the probability of that same shock under a distribution with thinner tails. Therefore, investors would give a negative utility to portfolio returns that come from a distribution with fat tails. In Figure 6.2 we have plotted the combinations of sample standard deviations and sample kurtosis as functions of the hedge ratio. Just as in Figure 6.1 the dots represent the hedge ratios and we have chosen $w = 0.5$. The sample kurtosis is computed by

$$\kappa \equiv \frac{\sum_{t=0}^{T-1} (r_{t+1}^h - \bar{r}_{t+1}^h)^4}{V(r_{t+1}^h)^2},$$

where \bar{r}_{t+1}^h is the sample mean and $V(r_{t+1}^h)$ the sample variance of r_{t+1}^h . The lowest dots correspond to the highest hedge ratios again. Interestingly, we see that a higher hedge ratio leads to a lower standard deviation and a higher kurtosis. In other words, a trade-off exists between these two measures of risk.

In Table 6.4 we show results on the optimal mean-variance and minimum-variance hedge ratios. We have chosen to set the risk aversion parameter λ equal to 1. We also report results on the unhedged portfolio ($h = 0$). From the table we see that the hedge ratios are all close to one. Strikingly, the standard deviations on the hedge ratios are large. The hedged portfolio means, variances and kurtosis do not differ much from each other. Shifting

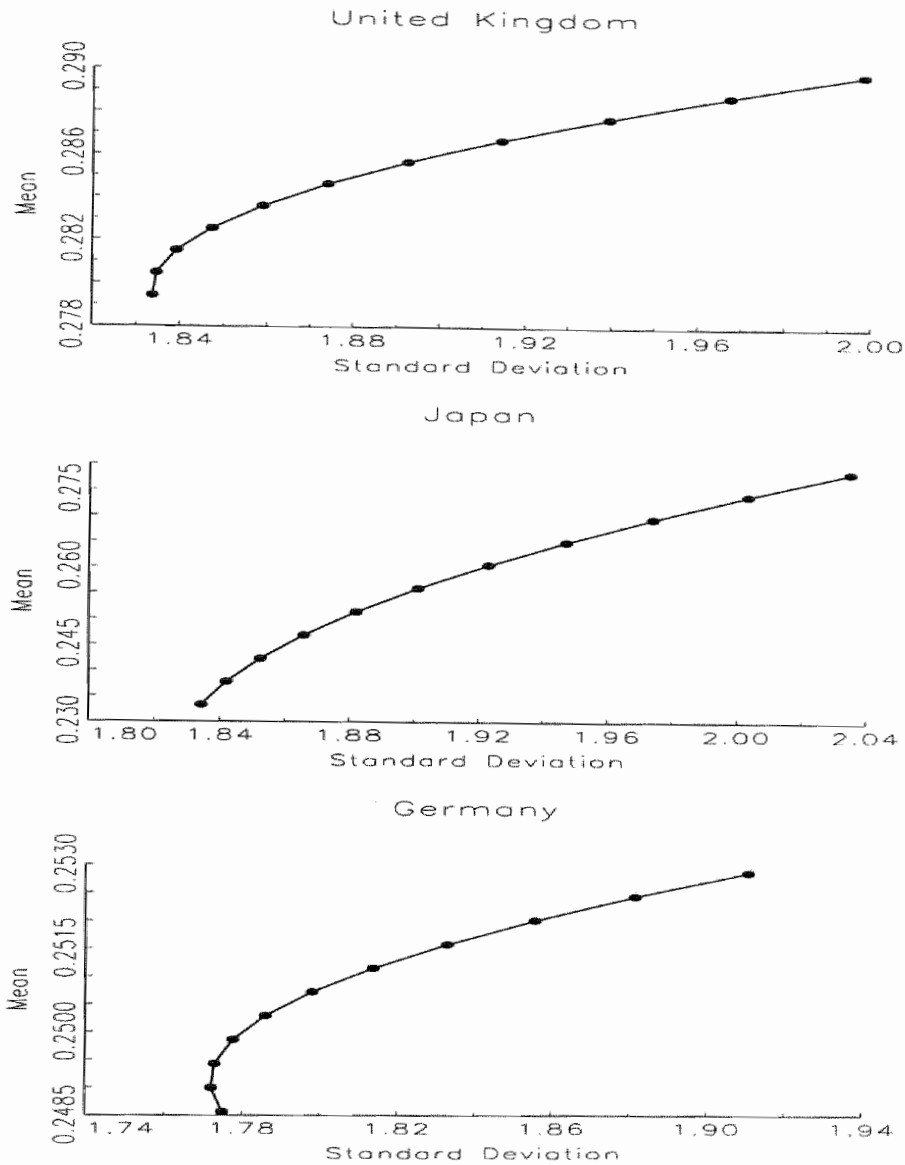


FIGURE 6.1: Means and Standard Deviations of r_{t+1}^h
Hedge ratios range from 0 to 1 with steps of 0.1. In every panel the lowest dot corresponds with $h=1$.

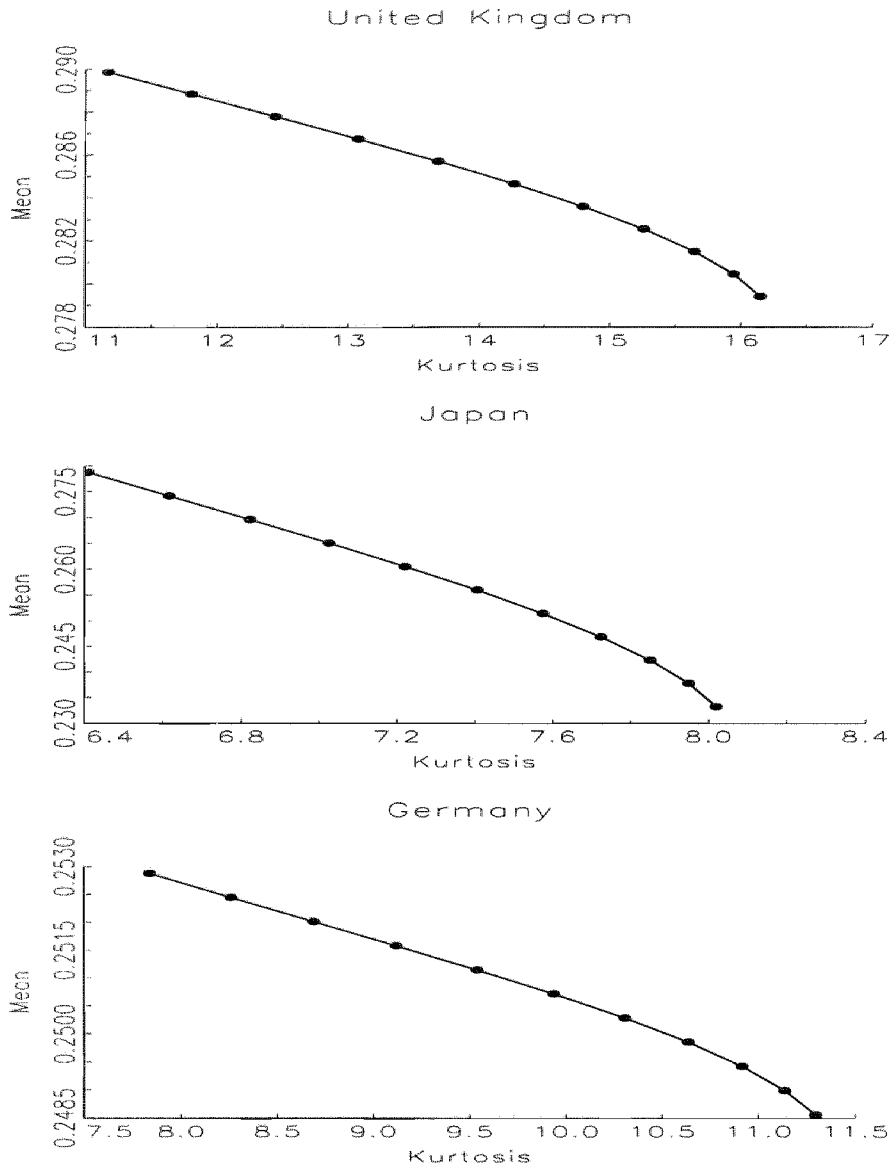


FIGURE 6.2: Standard Deviations and Kurtosis of r_{t+1}^h
Hedge ratios range from 0 to 1 with steps of 0.1. In every panel the lowest dot corresponds with $h=1$.

TABLE 6.4: Portfolio Characteristics of Hedged Portfolios r_{t+1}^h

| | $w=0.5$ | | | $w=0.8$ | | |
|-----------------|---------|---------|---------|---------|---------|---------|
| | UK | JP | GE | UK | JP | GE |
| \hat{h}_{mnv} | 0.946 | 1.051 | 0.863 | 1.059 | 1.176 | 0.997 |
| | (1.663) | (1.826) | (1.628) | (2.119) | (2.301) | (2.088) |
| μ_{mnv} | 0.280 | 0.231 | 0.249 | 0.278 | 0.256 | 0.266 |
| σ_{mnv} | 1.834 | 1.832 | 1.772 | 1.906 | 1.869 | 1.864 |
| κ_{mnv} | 16.05 | 8.043 | 11.06 | 12.32 | 9.827 | 10.61 |
| \hat{h}_{miv} | 0.975 | 1.202 | 0.874 | 1.077 | 1.273 | 1.005 |
| | (1.663) | (1.821) | (1.629) | (4.240) | (4.601) | (4.176) |
| μ_{miv} | 0.280 | 0.224 | 0.249 | 0.278 | 0.254 | 0.266 |
| σ_{miv} | 1.834 | 1.828 | 1.772 | 1.906 | 1.868 | 1.864 |
| κ_{miv} | 16.11 | 8.063 | 11.09 | 12.32 | 9.818 | 10.61 |
| $\mu_{h=0}$ | 0.290 | 0.279 | 0.253 | 0.278 | 0.282 | 0.267 |
| $\sigma_{h=0}$ | 1.998 | 2.035 | 1.911 | 1.906 | 1.938 | 1.893 |
| $\kappa_{h=0}$ | 11.17 | 6.407 | 7.838 | 11.32 | 9.485 | 9.821 |

Notes: Weekly data; sample period January 2, 1980 – May 31, 1995. Hedge ratios are computed for an American investor. The subscripts 'mnv', 'miv' and $h=0$ refer to the mean-variance, minimum-variance and zero-hedge hedge strategies, respectively. w is the weight put on American stock index investments. μ, σ and κ are the sample mean, standard deviation and kurtosis, respectively.

the portfolio weight towards home investments gives similar results.

The main question that we may ask at this point is how these results would change when we explicitly allow the underlying distributions to be fat-tailed. At this point several ideas may be exploited. In the optimization of the mean-variance expected utility function we found that the correlations between the unhedged portfolio return and the payoffs on the hedge instruments are an important ingredient in the hedge ratios. Consequently, we need a multivariate density function to describe all the returns. The multivariate normal distribution is easy to work with, since it is completely defined by its mean vector and covariance matrix. Suitable multivariate densities with fat tails are hard to find, however. For example, the multivariate Student t allows only one degrees of freedom parameter, which governs the overall leptokurtosis (Johnson and Kotz, 1972). By looking at the data we find that this is an undesirable characteristic, since we found different

levels of kurtosis for each of the stock index returns, exchange rate returns and interest rates. The same objection applies to the broader class of elliptical distributions (Owen and Rabinovitch, 1983). These distributions have a kernel that is based on the quadratic form $(x-\mu)'\Omega^{-1}(x-\mu)$, with μ and Ω a location vector and a scale matrix, respectively. Members of this class include, among others, the multivariate normal, multivariate Student t , and multivariate symmetric stable distributions. Although elliptical distributions have some desirable characteristics regarding portfolio theory we have opted for a somewhat different approach.¹¹

The expected utility function of the investor is a function of the hedged portfolio return, which in itself is a function of the underlying financial instruments that make up the portfolio (see (6.4)). Rather than specifying a multivariate distribution on the vector of portfolio elements, we specify a distribution directly on r_{t+1}^h : $f(r_{t+1}^h|h, w; \theta)$, with θ containing the distribution parameters. We condition on the choice variables (h and w) of the investor. The advantage is that f is a univariate distribution whose parameters can be easily estimated from the data. Hedge ratios can be found in the usual way by solving the first-order conditions of the expected utility function.

In this chapter we have chosen for the Student t distribution on r_{t+1}^h , although the exercises below can also be done for other distributions. The Student t is a continuous mixture of univariate normal distributions. The mixing variable is based on the chi-square distribution. The functional form of the distribution that we use here is

$$f(r_{t+1}^h|h, w; \theta) = \frac{\Gamma((1+q)/2)q^{q/2}\sqrt{V}}{\Gamma(1/2)\Gamma(q/2)} [q + V(r_{t+1}^h - m)^2]^{-(q+1)/2}$$

The parameter vector θ is given by the vector (q, m, V) . The parameter q is the degree of freedom parameter. Moments of order r of the Student t exist only when $r < q$. The parameter m is a location parameter and V is a scale variable. If $q \rightarrow \infty$, $f(r_{t+1}^h|h, w; \theta)$ approaches the normal distribution. Mean and variance are given by $E r_{t+1}^h = m$ and $\text{Var}(r_{t+1}^h) = V^{-1}q/(q+1)$, respectively. Conditional on the choice variables of the investor (h, w), the parameter vector θ can be estimated by maximizing the logarithmic likelihood function

$$\ln L(r^h|h, w; \theta) = \sum_{t=0}^{T-1} \ln f(r_{t+1}^h|h, w; \theta)$$

The algorithm we use to find the hedge ratios is given by the following steps. We take a sequence of hedge ratios in the range $[0, 1]$. For each of the hedge ratios we compute

¹¹See Owen and Rabinovitch (1983) and Ingersoll (1987) for portfolio issues under elliptically distributed random variables. A main feature of this class of distributions is that the CAPM can be extended to non-normal distributions.

the parameters of the distribution $f(r_{t+1}^h|h, w; \theta)$ by maximum likelihood. We use the parameters to compute the moments in the expected utility function. This gives us a sequence of expected utility values as function of the hedge ratio. Finally, we pick out the hedge ratio for which we find the highest expected utility.

The utility functions we chose include the mean-variance and minimum-variance utility functions that we used before. The difference is that we use the estimated moments from the Student t now instead of the sample moments that we applied earlier. These two utility functions do not explicitly take into account the excess kurtosis found in the data. We therefore chose a utility function that is an extension of the minimum variance utility function. This utility function measures risk by adding variance and kurtosis

$$E(r_{t+1}^h) = -\text{Var}(r_{t+1}^h) - \gamma \text{Kurt}(r_{t+1}^h)$$

where

$$\begin{aligned}\text{Var}(x) &= E(x - Ex)^2 \\ \text{Kurt}(x) &= \frac{E(x - Ex)^4}{\text{Var}(x)^2}\end{aligned}$$

By shifting γ an investor can put more or less weight on the kurtosis. We also investigated the implications of a utility function that minimizes kurtosis alone. The moments Ex^j , ($j = 1, \dots, 4$) were computed numerically.

Including kurtosis into the risk evaluation of an investor is a sensible action. When kurtosis is included in the utility function more weight is put on the tails of the distribution. Consequently, hedge ratios will be chosen in such a way that the effect of extreme observations on the portfolio is taken care of. Another way of applying this idea is put forward in the literature on extreme values and portfolio choice (see Jansen, Koedijk, and De Vries, 1994 and the references therein).

In Table 6.5 we present some results on finding hedge ratios under a Student- t distribution. In order to save space we only report results on $w = 0.5$. The results for $w = 0.8$ do not differ substantially. The values for q show that the leptokurtosis in r_{t+1}^h is picked up by the Student t specification. In fact the likelihood values are significantly higher than the likelihoods for a normal distribution. We illustrate the fit of the Student t in Figure 6.3. For comparison we also plotted the corresponding normal distribution. The sample frequencies were computed for a total number of 100 intervals. The Student t is clearly better at fitting the observations close to zero. Likewise we see that there is a considerable number of observations at the left side of the empirical distribution. This is picked up better by a fat-tailed distribution like the Student t . The hedge ratios found in Table 6.5

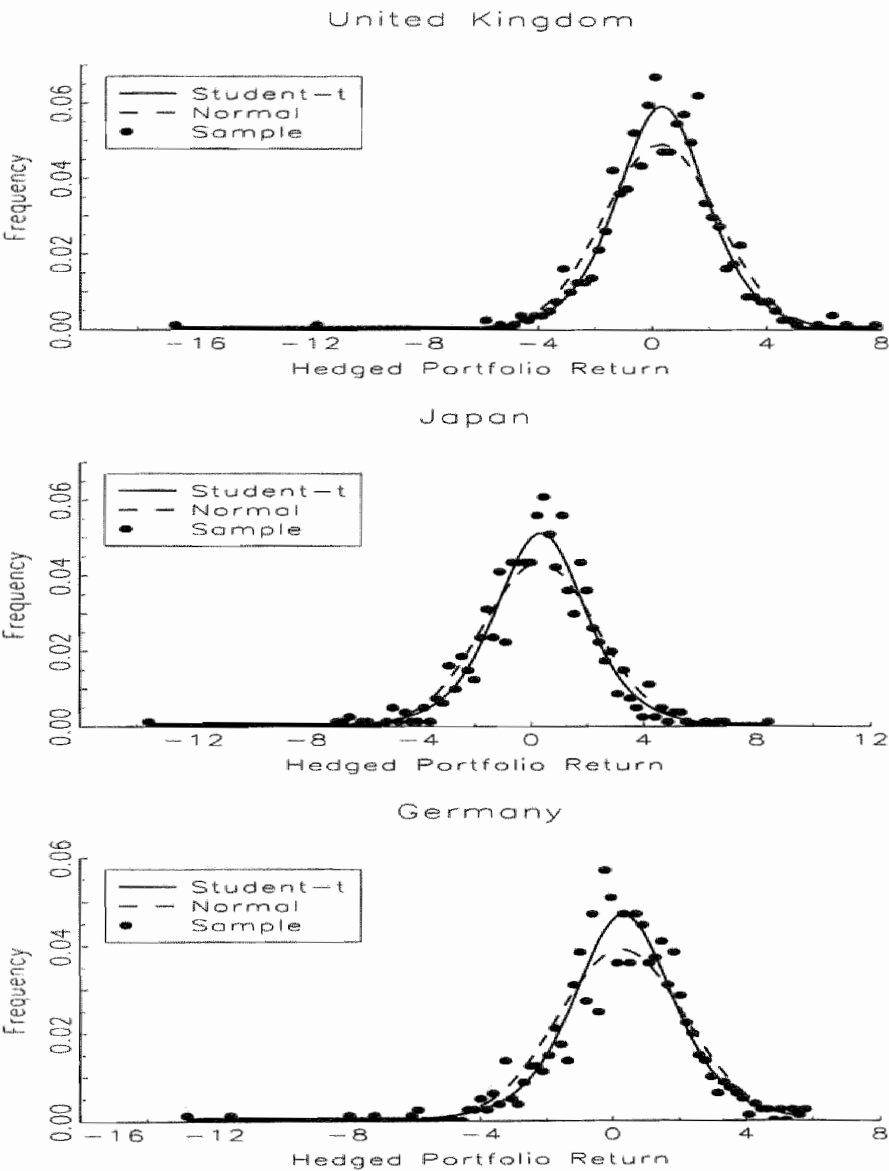


FIGURE 6.3: Expected and Sampled Frequencies of Fully Hedged Returns

TABLE 6.5: Hedged Portfolios: Student- t

| | mean-variance | | | minimum variance | | |
|-----------|------------------|------------------|------------------|------------------|------------------|------------------|
| | UK | JP | GE | UK | JP | GE |
| \hat{h} | 1 | 0.3 | 1 | 1 | 1 | 0.9 |
| m | 0.343 (0.057) | 0.321 (0.062) | 0.354 (0.055) | 0.343 (0.057) | 0.304 (0.056) | 0.351 (0.056) |
| V | 0.514 (0.041) | 0.430 (0.037) | 0.542 (0.043) | 0.514 (0.041) | 0.563 (0.050) | 0.542 (0.043) |
| q | 5.460 (0.931) | 5.217 (0.943) | 5.332 (0.891) | 5.460 (0.931) | 4.134 (0.629) | 5.346 (0.895) |
| μ | 0.279 | 0.265 | 0.249 | 0.279 | 0.233 | 0.249 |
| σ | 1.834 | 1.947 | 1.775 | 1.834 | 1.834 | 1.772 |
| κ | 16.15 | 7.024 | 11.29 | 16.15 | 8.019 | 11.14 |

| | variance-kurtosis | | | minimum kurtosis | | |
|-----------|-------------------|------------------|------------------|------------------|------------------|------------------|
| | UK | JP | GE | UK | JP | GE |
| \hat{h} | 1 | 0.4 | 1 | 0 | 0 | 0 |
| m | 0.343 (0.057) | 0.320 (0.061) | 0.354 (0.055) | 0.336 (0.064) | 0.320 (0.066) | 0.322 (0.062) |
| V | 0.514 (0.041) | 0.452 (0.039) | 0.542 (0.043) | 0.397 (0.031) | 0.367 (0.031) | 0.433 (0.035) |
| q | 5.460 (0.931) | 4.997 (0.876) | 5.332 (0.891) | 6.011 (0.741) | 5.967 (1.194) | 5.733 (1.058) |
| μ | 0.279 | 0.261 | 0.249 | 0.289 | 0.279 | 0.253 |
| σ | 1.834 | 1.923 | 1.775 | 1.998 | 2.035 | 1.911 |
| κ | 16.15 | 7.220 | 11.29 | 11.17 | 6.407 | 7.838 |

Notes: Weekly data; sample period January 2, 1980 – May 31, 1995. Hedge ratios are computed for an American investor that invests 50% at home and 50 % abroad. w is the weight put on American stock index investments. μ, σ and κ are the sample mean, standard deviation and kurtosis of the hedged returns $r_{t+1}^{\hat{h}}$, respectively.

for the minimum variance and the mean-variance utility functions do not differ much from what we found earlier. Thus, in that case allowing for fat-tailed distributions does not add much to the conventional analysis. This changes when kurtosis is explicitly included in the utility function of the investor. In the minimum kurtosis case hedge ratios change from values near one to zero, in accordance with the findings in Figure 6.2. The hedge ratios

under the variance-kurtosis utility function do not differ much from the mean-variance case. Only the hedge ratio for Japan changes slightly. The portfolio statistics are in line with the moments included in the utility functions. For example, the sample kurtosis is lower when we use a minimum-kurtosis utility function to compute hedge ratios.

6.3 Conclusions

In this chapter we have discussed some aspects of hedging foreign currency exposure. Typically, an investor maximizes an expected utility function with respect to the hedge ratios. We have put ourself the question in what sense the computation of optimal hedge ratios changes when we allow financial asset returns to be fat-tailed. We have seen that hedge ratios computed under a Student t distribution differ only from conventional hedge ratios when kurtosis is explicitly incorporated into the utility function. Consequently, we conclude that the risk profile of the investor is of equal importance as the underlying distributions of financial returns. This conclusion could be useful too in dynamic hedge models, where hedge ratios are based on time-varying conditional moments. Empirical results have shown that residuals from a GARCH model with a conditional normal error distribution still show some leptokurtosis in many cases. If the GARCH model is subsequently estimated with a conditional fat-tailed distribution for the returns, the hedge ratios will probably reflect the leptokurtosis better when the utility function values this leptokurtosis explicitly.

Chapter 7

A Factor Approximation Model for the International Arbitrage Pricing Theory

Applicability of international asset pricing models has been hampered by the numeraire dependence of these models. For example, the lack of success of the International Capital Asset Pricing Model, developed by Solnik (1974) and Adler and Dumas (1983), is partly due to the fact that the market portfolio is not efficient for investors denominating returns in different currencies. Testable restrictions cannot be formed in such a case. In contrast, the success of the International Arbitrage Pricing Theory can be mainly attributed to its numeraire independence. This independence is obtained by imposing severe restrictions on the factor structure of the model. Solnik (1983) derives a riskless portfolio by assuming that currency fluctuations have the same factor structure as asset returns expressed in numeraire terms. Ikeda (1991) implicitly assumes that exactly one currency factor per country exists.

In this chapter we develop a model that poses less restrictions on the currency factor structure. This can be done by using a modified version of the numeraire-free principal components method of Koedijk and Schotman (1989) together with the factor approximation results from Chamberlain and Rothschild (1983). Our model allows an analysis of exchange rate returns simultaneously with asset returns.

The outline of this chapter is as follows. Section 7.1 outlines the traditional international arbitrage pricing theories and focuses on their limiting assumptions. In Section 7.2 our general framework is presented. First we prove that arbitrage portfolios can be formed in case of a numeraire independent factor structure. Then we show how factor structures can

be formed independent of the numeraire currency chosen. Finally we use the results from Chamberlain and Rothschild (1983) to complete the model. Section 7.3 contains some concluding remarks.

7.1 International Arbitrage Pricing Theory

The Arbitrage Pricing Theory (APT) has gained wide acceptance as an asset pricing model since its development by Ross (1976). An important reason for this is that only a limited number of assumptions are needed for the APT to be valid (Jarrow, 1988). The most important assumption is that asset returns can be described by a factor model. Typically, a cross section of a large number of assets shows considerable co-movements. The sources of these common movements are not known in general. The APT tries to capture fluctuations in asset returns by specifying a statistical factor model for unexpected asset returns. Subsequently, the APT states that, under suitable circumstances, the vector of expected returns is spanned by a vector of ones and a number of vectors containing the factor loadings.

In an international context several models have been suggested for the description of asset returns. However, many models turn out to be restrictive with respect to investor behaviour. A well-studied model is the International Asset Pricing Model (IAPM), which is an international variant of the well-known Capital Asset Pricing Model (CAPM). As Adler and Dumas (1983) show, however, the international market portfolio is not efficient for investors calculating their returns in different numeraires. This has led researchers to develop other international asset pricing models, of which the International Asset Pricing Theory (IAPT) seems to be the most promising. In this section we will concentrate on the latter model.

Consider a world of N countries. In every country there exists one risky asset and one locally riskless bond. The random return on the risky asset in country i and currency j is r_{ij}^t . We model the local return in country i as the sum of expected local return, a linear combination of K factors and an idiosyncratic return term

$$r_{ij}^t = E r_{ij}^t + \sum_{k=1}^K b_{ik} f_k + \epsilon_i. \quad (7.1)$$

The factors f_k are international factors, *i.e.* they do not depend on the country index i . We assume that they have zero mean: $E f_k = 0$ ($k = 1, \dots, K$). The factors need not be orthogonal, but can be made so by rotating the factors appropriately. Note that this can be done without loss of generality, because we have put no limiting restrictions on the

factors f_k and the corresponding factor loadings b_{ik} . The sensitivities or factor loadings b_{ik} do depend on the country index. The random term ϵ_i is also called idiosyncratic return and has mean zero and is uncorrelated with the factors f_k . In order for the APT to hold we have to assume that N is large.² This ensures that residual risk, measured by the variance of the idiosyncratic return, tends to zero.

We assume that currency returns are driven by news, *i.e.* currency changes are unpredictable and exchange rates are a random walk. Let s_i^j be the random currency return measured as the price of currency i in terms of currency j . We have

$$s_i^j = \nu_i^j,$$

where ν_i^j is an independently and identically distributed random variable. In the original IAPT model of Solnik (1983) it is assumed that currency fluctuations obey the same factor model as local asset returns do

$$\nu_i^j = \sum_{k=1}^K c_{ik}^j f_k + \eta_i^j,$$

with η_i^j an idiosyncratic term and c_{ik}^j the currency factor loadings. When we express the return on a risky asset of country i in any other currency, say j , we obtain

$$\begin{aligned} r_i^j &= r_i^i + s_i^j \\ &= Er_i^j + \sum_{k=1}^K (b_{ik} + c_{ik}^j) f_k + \epsilon_i + \eta_i^j, \end{aligned}$$

which is again a factor model. Note that we have $Er_i^i = Er_i^j$, because we assumed that exchange rate returns are induced by unpredictable news. The new factor loadings $b_{ik} + c_{ik}^j$ and the new idiosyncratic terms $\epsilon_i + \eta_i^j$ now depend on the numeraire currency chosen. Solnik (1983) assumes that the idiosyncratic risk approaches zero when the number of countries goes to infinity ($N \rightarrow \infty$). The factor structure, however, remains the same whatever currency we choose to denominate our returns in. We can apply the usual APT arbitrage arguments to get the vector pricing relationship

$$Er^j = \lambda_0^j \iota_N + \sum_{k=1}^K \lambda_k^j (b_k + c_k^j), \quad j = 1, \dots, N,$$

where Er^j , ι_N , b_k and c_k^j are vectors with length N . For all numeraire currencies j , every element in this vector equation represents the pricing relationship for a particular country i ($i = 1, \dots, N$). The λ_k^j ($k = 0, \dots, K$; $j = 1, \dots, N$) are scalars that can be regarded as

²This differs from the domestic case in which we need a large amount of domestic assets. However, Ikeda (1991) notes that the IAPT also holds under the assumption of a limited number of countries each with a large amount of risky assets. See also Solnik (1983).

measures of combined risk, since they incorporate the non-diversifiable exchange rate and asset return risk.

Ikeda (1991) notes that the restriction that exchange rate returns should follow the same factor structure as asset returns is quite strong. If currency returns are governed by a totally different factor structure than asset returns the idiosyncratic risk – represented by the variance-covariance matrix of ϵ_i – could become arbitrarily large. This would lead to a worsening of the fit of the arbitrage pricing relationships (see Conner and Korajczyk, 1993). When we leave out the currency factor restriction we obtain the factor model

$$\begin{aligned} r_i^j &= r_i^i + s_i^j \\ &= Er_i^j + \sum_{k=1}^K b_{ik} f_k + \epsilon_i + \nu_i^j. \end{aligned}$$

Remember that $Er_i^i = Er_i^j$. The currency news term ν_i^j represents the currency risk that investors from country j have. Unlike the residual risk η_i^j the currency risk ν_i^j cannot be diversified away. Furthermore, ν_i^j depends on the asset number i implying that an arbitrage portfolio cannot be constructed without additional assumptions.³

It is possible to formulate an arbitrage portfolio in the Ikeda (1991) model by using the locally riskless bonds in every country. By offsetting the investment in the risky asset of country i with an equal short position in the bond, the currency exposure that the investor has in country i becomes equal to zero. The arbitrage portfolio now consists of the following parts. Let $w = (w_1, \dots, w_N)$ be portfolio weights for the investment in the risky assets. The arbitrage portfolio for the risky assets is

$$\begin{aligned} w' b_k &= 0, & k = 1, \dots, K, \\ w' \epsilon &\approx 0, \end{aligned}$$

where $b_k = (b_{1k}, \dots, b_{Nk})$ and $\epsilon = (\epsilon_1, \dots, \epsilon_N)$. The investment in risky asset i , w_i , is offset by an investment $-w_i$ in the riskless bond of country i . The currency j value of the return on the bond is given by

$$R_i^j = R_i^i + \nu_i^j, \quad i = 1, \dots, N,$$

with R_i^i the local riskless return on the bond, and R_i^j its random counterpart when denominated in currency j . Applying this arbitrage portfolio to all risky and riskless assets available leads to

$$r_\theta^j \equiv \sum_{i=1}^N w_i r_i^j - \sum_{i=1}^N w_i R_i^j$$

³The factor structure for exchange rates that was presented in Chapter 4 has the same problems. The number of factors (or news components) is equal to the number of countries N . In case of a limited number of countries and a large number of risky assets per country arbitrage portfolios can be formulated. We will not pursue this issue here.

$$\begin{aligned} &= \left(w'Er^j + \sum_{k=1}^K w'b_kf_k + w'\epsilon + w'\nu^j \right) - (w'R + w'\nu^j) \\ &= w'(Er^j - R), \end{aligned}$$

with $R = (R_1^1, \dots, R_N^N)$. This portfolio return does not incorporate any exchange rate components anymore. The return on the arbitrage portfolio is r_θ^j . We can now apply the arbitrage argument again. The return on the arbitrage portfolio should be equal to zero: $w'(Er^j - R) = 0$. This leads to the pricing relationship

$$Er^j = R + \sum_{k=1}^K \bar{\lambda}_k^j b_k, \quad j = 1, \dots, N.$$

This equation shows us that the expected returns in currency j corrected for the riskless bond are a linear combination of the factor loadings. The factor risk prices do not depend on currency j : the hedged securities are priced in terms of local factor risk prices.

The IAPT model of Solnik (1983) has been empirically investigated by Cho, Eun and Senbet (1986). Testing the IAPT is much easier than testing the IAPM of Adler and Dumas (1983), due to the numeraire invariant factors in Solnik's IAPT. Furthermore, a world market portfolio is not needed. A first step in any testing procedure of the IAPT (or APT in general) is to estimate a number of factors. The APT itself does not give us any guidelines for this, because it is a statistical model. The search for the number of factors is influenced by the assumptions that the investor makes with respect to the residual terms of the factor model.

In the derivation of the IAPT models of Solnik (1983) and Ikeda (1991) we have not specified any structure on the covariance matrix of the idiosyncratic returns. In a strict factor model the idiosyncratic terms should be uncorrelated with each other, implying that the covariance matrix is diagonal (Basilevsky, 1994). However, this could lead to undesirable features in the factor structure. Some countries in the sample could have a common component that is not important for the pervasive sources of uncertainty in the total system of countries. For example, the returns of risky assets in European countries could be driven by a typical European component, which is not important for other countries in the sample. In order for the strict factor model to be valid the "European" factor should be included in the set of pervasive factors for the whole sample. This could lead to a proliferation of factors. Thus, it seems warranted to search for a model that has a flexible factor structure with respect to asset and exchange rate returns. In the next section we will show that this is indeed possible using approximate factor structures.

7.2 Factor Approximation

Chamberlain and Rothschild (1983) have relaxed the assumptions of the original APT by allowing the covariance matrix of the idiosyncratic returns to be a general positive definite matrix, *i.e.* the residual terms can be correlated. They show that the APT remains valid, if the amount of covariance is restricted. This model is called an approximate factor model. In this section we will describe shortly the techniques that underlie factor approximation. It appears that it is closely related to principal component analysis. We will use the results from Chamberlain and Rothschild (1983) to design an international version of the APT that allows a flexible exchange rate factor structure and at the same time retains the numeraire independence assumption of the IAPT. To this end we use a special version of principal component analysis that was developed by Koedijk and Schotman (1989).

To explain the importance of approximate factor structures look at the general K -factor model

$$y^N = c^N + B^N f + \epsilon^N,$$

where y^N is a random vector of data and c^N is a constant vector. The matrix B^N with columns b_1^N, \dots, b_K^N contains the factor loadings on the K unknown factors f . The factors have mean zero and are assumed to be rotated in such a way that $E f f' = I_K$. The superscripts mark the length of the vectors. The idiosyncratic vector ϵ^N has mean zero, covariance matrix V^N , and is uncorrelated with the factors. As mentioned above this covariance matrix has to be diagonal in a strict factor model. In an approximate factor model Chamberlain and Rothschild (1983) assume that

$$\begin{aligned} \lim_{N \rightarrow \infty} \frac{1}{N} B^N B^{N'} &= B^*, \\ \lim_{N \rightarrow \infty} \|V^N\|_2 &\leq M < \infty, \end{aligned}$$

with B^* a nonsingular matrix, and M a positive scalar. The covariance matrix V^N can have a general structure now under the restriction that it satisfies the second condition. This condition says that the eigenvalues of V_N should be bounded when we let the number of elements in the vectors go to infinity ($N \rightarrow \infty$). The matrix norm $\|\cdot\|_2$ is the spectral norm of a matrix, which is equal to the largest eigenvalue for symmetric positive semidefinite matrices. The covariance matrix Σ^N of the y^N vector is equal to

$$\Sigma^N = B^N B^{N'} + V^N.$$

The matrix $B^N B^{N'}$ has rank K . Chamberlain and Rothschild (1983) show that with increasing N the smallest eigenvalue of the variance of the pervasive components increases at the rate N . Intuitively, this must be the case since an additional element in the sample

(increasing N) has a loading on the pervasive risk plus an idiosyncratic component that can be diversified away. Consequently, the $(K+1)$ th eigenvalue – related to V^N – is assumed to be bounded in the approximate factor model. Furthermore, the approximate factor structure is unique with the assumptions made above.

The relationship between an approximate factor model and the eigenvalues of the covariance matrix Σ^N allows a principal components analysis to find the approximate factors. A principal component analysis is a linear transformation of a set of random variables to an orthogonal set which reproduces the original variance-covariance structure. Many linear transformations of the original data are possible. Here we are interested in a transformation that maximizes the variance of the original random variables in a sequential way. This is equivalent to solving for the eigenvalues of the variance-covariance matrix Σ^N (Basilevsky, 1994).

In general, three conditions have to be met in order to transform the original data matrix Y^N , with dimension T by N , into its principal components matrix Z^N :

$$\begin{aligned} Z^N &= Y^N Q B^N, \\ Z^{N'} Z^N &= \Lambda^N, \\ B^{N'} Q B^N &= I_N. \end{aligned}$$

In the first condition, the scaling matrix Q is positive definite symmetric and B^N non-singular. The matrix B^N is a matrix containing loadings on the principal components. In the second line, Λ^N is a diagonal matrix. The elements of Λ^N are in descending order of magnitude: $\lambda_1^N \geq \dots \geq \lambda_N^N$. Since they determine the variance of the principal components, the first principal component explains most of the variance. I_N is an identity matrix of size N . The third line is a normalization. The corresponding eigenvalue problem is to find the λ_i^N ($i=1, \dots, N$) from

$$\left(Q^{\frac{1}{2}} Y^{N'} Y^N Q^{\frac{1}{2}} \right) \left(Q^{\frac{1}{2}} B^N \right) = \left(Q^{\frac{1}{2}} B^N \right) \Lambda^N,$$

where $Q^{\frac{1}{2}} Q^{\frac{1}{2}} = Q$. The principal components Z^N depend on the choice of Q , which is not determined in the principal components analysis, but must be specified *a priori*. In conventional principal components analysis, Q is usually diagonal with elements $q_{ii} = 1/\hat{\sigma}_i^2$ ($i = 1, \dots, N$), where $\hat{\sigma}_i^2$ is the sample variance of column i of Y^N . Once Q has been specified, one and only one matrix B^N of principal components loadings exists, which satisfies the three basic conditions. In other words, the transformation from Y^N to Z^N is unique.

The principal components analysis described in the previous paragraph was used by Koedijk

and Schotman (1989).⁴ The main innovation of their method lies in the choice of the scale matrix Q . Instead of using Q as a normalization of the data matrix Y^N , it is chosen in such a way that the principal components matrix Z^N is independent of a common numeraire currency. Remember that this is a vital ingredient in the derivation of a general IAPT.

Let the random vector r^j contain the risky assets for the N countries denominated in currency j . Similarly let s^j be an $(N-1)$ -vector with currency returns, where we have skipped the element $s^j_j = 0$. Our data matrix Y^N then consists of $2N-1$ columns of which the first $N-1$ columns are currency returns and the last N columns are returns on riskless assets in some numeraire. We are interested in finding a matrix P that satisfies the following equation

$$\begin{pmatrix} s^j \\ r^j \end{pmatrix} = P \begin{pmatrix} s^m \\ r^m \end{pmatrix}.$$

The matrix P $((2N-1) \times (2N-1))$ governs the currency translation from currency m to currency j . Remember that exchange rate returns obey the identity $s^j_m = -s^m_j$. The matrix P then becomes

$$P = \begin{pmatrix} -1 & 0' \\ -\iota_{2N-2} & I_{2N-2} \end{pmatrix}$$

where I_{2N-2} is an identity matrix of rank $2N-2$, and ι_{2N-2} is a $(2N-2)$ -vector of ones. It can be shown that the matrix P is unipotent: $PP = I_{2N-1}$. This is a necessary condition for P since applying the same transformation twice should lead to the original vector $(r^{j'}, s^{j'})'$.

The condition under which a change in the numeraire currency does not affect the principal components structure is based on the existence of a matrix Q such that $Q = P'QP$. Since Q should be positive definite symmetric, this condition can always be met (see Koedijk and Schotman, 1989).

The existence of an approximate factor structure in an environment with a large number of assets allows a principal components analysis. In an international context we have seen that the principal components can be made independent from the numeraire. Consequently, the restrictions that Solnik (1983) and Ikeda (1991) put on the currency structure of their IAPT models can be significantly weakened when use is made of a numeraire independent principal components analysis. A heuristic statistic for determining the number of dominant components is given in Koedijk and Schotman (1989). This statistic is based on the eigenvalues of the principal components analysis λ_i^N ($i = 1, \dots, N$) and measures the percentage of variation that is taken out of the sample by the first k principal components.⁵

⁴Koedijk and Schotman (1989) perform the principal component analysis on a set of real exchange rates. We will extend the universe of assets to include asset returns as well.

⁵Remember that the principal component matrix Z^N in an IAPT setting consists of $2N-1$ components.

7.3 Conclusions

In this chapter we have developed an extension to the international arbitrage pricing theory. Using the approximation of a factor structure in asset returns with principal components, we have demonstrated that a numeraire independent international arbitrage pricing theory can be derived. Our new model involves a numeraire free type of principal components analysis. Compared to Solnik (1983) and Ikeda (1991) the contribution of our model is that it puts no limits on the factor structure of currency returns.

Chapter 8

Conclusions

Volatility is an important concept in financial economics. At an intuitive level volatility is something that investors do not like. Volatility makes investors uncertain of what decisions they should make. Consequently, it becomes important to measure volatility in some way and incorporate these measurements in the investors' decision process. In this thesis we have discussed several ways to measure volatility and how to deal with it in certain decision problems. In the empirical financial literature volatility models abound, and this thesis does not try to encompass all of them. Instead, we focussed on a subset of volatility models and from this subset we have chosen some interesting empirical examples. Nevertheless, the models that we have discussed in this thesis have important content for future volatility modelling and decision making in a risky world. In particular, stochastic volatility models will probably be used more and more once the merits of the several estimation techniques have been crystalized out. Ongoing increases in computer power will accomodate this process. Topics in finance that rely heavily on continuous time arguments, like contingent claims pricing and dynamic portfolio allocation, can be analyzed using the stochastic volatility models. In this chapter we will restate some of the conclusions that arose from the chapters of this thesis and put them into perspective with respect to future research.

Volatility becomes useful once we are able to measure it. In Chapter 2 we have presented a flexible framework for volatility modelling that is based on measurability concepts. Basically, we specified an information structure that contains sufficient information for volatility models to exist. We have shown that two popular volatility model classes are contained within the framework. The first model class comprises AutoRegressive Conditional Heteroskedasticity (ARCH) models and all its colorful descendants. ARCH models are characterized by the fact that the volatility process is governed by (known) information from

the past. The second class of models is designated by the term Stochastic Volatility (SV) models. In SV models the volatility at some point in time is not only governed by information from the past, but also by a contemporaneous stochastic element. This makes the volatility a latent process. Stochastic volatility models have a long history in theoretical finance. Continuous time models of asset price behavior and portfolio selection all have a volatility component that reflects contemporaneous information. In fact, we show that a SV model naturally arises when we discretize a continuous time model.

The estimation of volatility processes is an important topic in empirical financial research. In this area a practical distinction between ARCH and SV models shows up. Since the information set of ARCH models consists only of information from the past, estimation by maximum likelihood is relatively simple. This is certainly not the case for SV models, for which the volatility process is latent. In Chapter 3 we have looked at the estimation of SV models. The analytical expression of the estimation problem is given by an awkward multidimensional integral. Several approaches have been suggested in the literature. We have chosen to work with a method that transforms the SV model to a conditional linear state-space model, for which we know how to obtain estimates for the latent (volatility) process. The shape of the measurement error noise is determined by a mixture of normal distributions with free mean and variance parameters. Estimates of the hyperparameters were obtained by simulation methods based on the Gibbs sampler. We performed our estimation procedure both in a classical framework through a simulated EM algorithm, and in a Bayesian framework in which the hyperparameters are considered to be random variables. We illustrated our estimation technique with an analysis of exchange rate news for weekly exchange rates between the four major currencies (U.S. dollar, Japanese yen, German mark and U.K. pound). Our sample period covers the whole floating rate period starting in 1973. An important conclusion is that the estimated volatility processes differ substantially from one estimation procedure to the next. Furthermore, the standard deviations on the volatilities are quite large. In part this may be due to the low signal-noise ratio in SV models. The results imply that extreme care has to be taken in using volatility processes in financial decision making processes, in particular the valuation of derivative securities.

The diagnostic Lagrange Multiplier statistics of the univariate SV models in Chapter 3 showed that the univariate SV model with an AR(1) volatility specification is probably misspecified in case of describing exchange rates. Also, significant spillover effects seem to be present in the exchange rates studied. This motivates multivariate models of exchange rate returns. In Chapter 4 we have tackled the problem of estimating multivariate time series that are subject to changing volatility. As we showed many multivariate models that have been used up to now in the exchange rate literature have the problem that

the form – and consequently the implications – of the model change when the currency of denomination changes. In our model this is not the case. We arrive at our result by assuming that exchange rate news can be decomposed in country-specific news series. In other words, we construct a factor model for any set of exchange rates in which the currency-specific news series act as factors. In a subsample analysis we show that the variance of our news components changes over time. This leads us to consider a multivariate SV model of our system of exchange rates. The factor structure of exchange rates allows a parsimonious representation of the exchange rate system. When we estimate the model with the four currencies we find that significant spillover effects exist indeed. It seems that U.S. dollar volatility spreads quickly over to the other volatilities in the system, whereas the volatilities from other countries do so in a much slower fashion. Also, it appears that the German mark is the least volatile of the major currencies over a period of almost 20 years. We also performed an analysis of exchange rate risk premia in our system of exchange rates. We find that risk premia fluctuate considerably over the sample period. Compared to previous studies our annualized risk premia are relatively high, reflecting the volatility estimation error that is explicitly taken care of in the computation of the risk premia.

Both ARCH models and SV models are statistical models, *i.e.* no economic theory has been proposed that explains their rationale. Nonetheless, on empirical grounds, both types of volatility models are successful in describing financial time series. A theory that tries to explain the emergence of time-varying volatility in financial markets is the Mixture of Distributions Hypothesis (MDH). In this theory the stochastic process of a financial asset is governed by a subordinated process that represents information arrival. A variable that is commonly assumed to approximate the amount of information arrival is trading volume. In Chapter 5 we present some empirical analyses on the role of volume within a SV model. We assume that volumes and volatility are determined at the same time and subsequently investigate the hypothesis that volume is a scale factor for volatility. The estimation results with the SV model confirm this hypothesis. The dataset for which we present results comprises daily data on returns and trading volumes of 5 individual stocks on the New York Stock Exchange and the S&P500 for the period January 1967 to December 1987. In the empirical literature on stock return time series considerable attention is given to the leverage effect. Positive and negative innovations might have a different impact on volatility. We also investigate the possibility of a leverage effect in our series. We find that it is indeed present in the S&P500 series, but for the individual stock return series the evidence appears to be mixed. It turns out that the leverage effect is mainly a tail phenomenon. Nevertheless, we repeated our estimation procedures for a modified SV model in which we allow correlation between the measurement error in the observation equation of the SV model on the one hand, and innovations in the volatility

equation on the other. As expected, this does not change the conclusions substantially: trading volumes remain a significant scaling factor for volatility.

In chapters 6 and 7 we investigate two important issues in international financial management. Chapter 6 is devoted to hedging foreign currency exposure in an internationally diversified portfolio. Currency hedging has gained a lot of attention in the literature. However, reconciling hedging strategies with the stylized fact of leptokurtosis in financial asset returns has not been analyzed until recently. We develop a model in which investors determine hedge ratios by maximizing an expected utility function in a one-period setting. The argument in the utility function is the return on an international portfolio that consists of local and foreign investments in a stock index, and forward contracts. We chose to model this total return by a Student t distribution. Not surprisingly, leptokurtosis is valued only in a utility function that incorporates higher moments explicitly. For a sample of weekly returns from 1980 to 1995 we find that an American investor would not hedge his bilateral currency exposure with respect to the United Kingdom, Japan and Germany, when he minimizes a risk measure based on kurtosis alone, whereas he would fully hedge the exposure under a minimum variance expected utility function. Thus, the investor's risk attitude should be tuned with respect to the shape of the relevant unconditional distributions that make up the total return on the portfolio.

In Chapter 7 we discussed the international allocation of wealth within an international version of the Arbitrage Pricing Theory. Any valid international asset pricing model should exhibit independence of the numeraire in which investors denominate their holdings of an internationally diversified portfolio. Otherwise, investors could make a riskless profit by switching from one numeraire to another. The International Arbitrage Pricing Theory (IAPT) is perhaps the most studied model in international finance. However, some strong restrictions on the currency parts of the model are needed in order for the model to exhibit independence from the numeraire. We propose another route, that allows a joint factor structure of asset and currency returns, which is independent of the numeraire chosen. The model is based on the observation that factors in an arbitrage pricing model may be found by doing a principal components analysis on the relevant time series in the sample. Under some conditions these approximate factors can be made independent of the numeraire by performing an appropriate rotation.

In this thesis we have tried to give some new insights on modelling volatility. At the same time it has reconfirmed old insights. It has shown that volatility is a difficult quantity to measure, both from a statistical and from an empirical viewpoint. Stochastic volatility results have shown that signal-noise ratios are very low. Furthermore, volatility processes turned out to be very volatile themselves. Many of the volatility models that exist today

are pure statistical models. No clear insights on the origins of volatility exist. Besides improving the statistical aspects of volatility models, it is vital for our understanding of financial markets to analyze the mechanisms that cause volatility in the first place. Microstructural studies seem to be a promising area in this respect. The advent of high frequency data on financial time series may facilitate this line of research. Another route that seems promising is the simulation of market settings in a laboratory. The quest continues.

Bibliography

- ABRAMOWITZ, M. AND N.C. STEGUN, (1970), *Handbook of Mathematical Functions*, Dover Publications, New York.
- ADLER, M. AND B. DUMAS, (1983), International Portfolio Choice and Corporation Finance: A Synthesis, *Journal of Finance*, **38**, 925–984.
- ADLER, M. AND B. PRASAD, (1992), On Universal Currency Hedges, *Journal of Financial and Quantitative Analysis*, **27**, 19–37.
- ADMATI, A.R. AND P. PFLEIDERER, (1988), A Theory of Intraday Patterns: Volume and Price Variability, *Review of Financial Studies*, **1**, 3–40.
- AKGIRAY, V. AND G.G. BOOTH, (1988), Mixed Diffusion-Jump Process Modeling of Exchange Rate Movements, *Review of Economics and Statistics*, **70**, 631–637.
- ANDERSEN, T.G., (1992), Volatility, working paper 144, Northwestern University.
- ANDERSEN, T.G., (1994a), Return Volatility and Trading Volume: an Information Flow Interpretation of Stochastic Volatility, unpublished manuscript, Northwestern University.
- ANDERSEN, T.G., (1994b), Stochastic Autoregressive Volatility: a Framework for Volatility Modeling, *Mathematical Finance*, **4**, 75–102.
- ANDERSEN, T.G. AND B.E. SØRENSEN, (1994), GMM Estimation of a Stochastic Volatility Model: A Monte Carlo Study, unpublished manuscript, Northwestern University.
- ANDERSON, R.W. AND J.-P. DANTHINE, (1981), Cross Hedging, *Journal of Political Economy*, **89**, 1182–1196.
- BAILLIE, R.T. AND T. BOLLERSLEV, (1989), The Message in Daily Exchange Rates: A Conditional Variance Tale, *Journal of Economic and Business Statistics*, **7**, 297–305.
- BAILLIE, R.T. AND T. BOLLERSLEV, (1990), A Multivariate Generalized ARCH Approach to Modeling Risk Premia in Forward Foreign Exchange Markets, *Journal of International Money and Finance*, **9**, 309–324.
- BAILLIE, R.T. AND T. BOLLERSLEV, (1991), Intra-day and Inter-market Volatility in Foreign Exchange Rates, *Review of Economic Studies*, **58**, 565–585.
- BAILLIE, R.T. AND R.J. MYERS, (1991), Bivariate GARCH Estimation of the Optimal Commodity Futures Hedge, *Journal of Applied Econometrics*, **6**, 109–124.
- BASILEVSKY, A., (1994), *Statistical Factor Analysis and Related Methods*, Wiley, New York.

- BAUER, R.M.M.J. AND F.G.M.C. NIEUWLAND, (1994), The Value of Embedding Proxies for Information Arrival in GARCH Processes, unpublished manuscript, RWTH Aachen.
- BEKAERT, G. AND R.J. HODRICK, (1993), On Biases in the Measurement of Foreign Exchange Risk Premiums, *Journal of International Money and Finance*, **12**, 115–138.
- BEKAERT, G., R.J. HODRICK, AND D. MARSHALL, (1993), The Implications of First Order Risk Aversion for Asset Market Risk Premiums, unpublished manuscript, Stanford University.
- BERGSTROM, A.R., (1990), *Continuous Time Econometric Modelling*, Oxford University Press.
- BERNSTEIN, P.L., (1992), *Capital Ideas: The Improbable Origins of Modern Wall Street*, The Free Press, New York.
- BLACK, F., (1976), Studies of Stock Price Volatility Changes, in *Proceedings of the 1976 Meetings of the American Statistical Association, Business and Economics Statistics Section*, 177–181.
- BLACK, F., (1990), Equilibrium Exchange Rate Hedging, *Journal of Finance*, **65**, 899–907.
- BLACK, F. AND M. SCHOLES, (1973), The Pricing of Options and Corporate Liabilities, *Journal of Political Economy*, **81**, 637–659.
- BOLLERSLEV, T., (1986), Generalized Autoregressive Conditional Heteroskedasticity, *Journal of Econometrics*, **31**, 307–328.
- BOLLERSLEV, T., (1988), A Conditionally Heteroskedastic Time Series Model for Speculative Prices and Rates of Return, *Review of Economics and Statistics*, **69**, 542–547.
- BOLLERSLEV, T., R.Y. CHOU, AND K.F. KRONER, (1992), ARCH Modeling in Finance: a Review of the Theory and Empirical Evidence, *Journal of Econometrics*, **52**, 5–59.
- BOLLERSLEV, T. AND I. DOMOWITZ, (1993), Trading Patterns and Prices in the Interbank Foreign Exchange Market, *Journal of Finance*, **48**, 1421–1443.
- BOLLERSLEV, T., R.F. ENGLE, AND D.B. NELSON, (1993), ARCH Models, forthcoming in *The Handbook of Econometrics*.
- BOLLERSLEV, T., R.F. ENGLE, AND J.M. WOOLDRIDGE, (1988), A Capital Asset Pricing Model with Time-varying Covariances, *Journal of Political Economy*, **96**, 116–131.
- BOOTHE, P. AND D. GLASSMAN, (1987), The Statistical Distribution of Foreign Exchange Rates: Empirical Evidence and Economic Implications, *Journal of International Economics*, **22**, 297–319.
- BRENNER, R.J., R.H. HARJES, AND K.F. KRONER, (1994), Another Look at Alternative Models of the Short-Term Interest Rate, unpublished manuscript, University of Arizona.
- BROCK, W.A. AND A. KLEIDON, (1992), Periodic Market Closure and Trading Volume: A Model of Intra Day Bids and Asks, *Journal of Economic Dynamics and Control*, **16**, 451–489.
- CAMPBELL, J.Y., S.J. GROSSMAN, AND J. WANG, (1993), Trading Volume and Serial Correlation in Stock Returns, *Quarterly Journal of Economics*, **101**, 905–939.

- CAMPBELL, J.Y. AND L. HENTSCHEL, (1992), No News is Good News: An Asymmetric Model of Changing Volatility in Stock Returns, *Journal of Financial Economics*, **31**, 281-318.
- CANOVA, F., (1993), Modelling and Forecasting Exchange Rates with a Bayesian Time-Varying Coefficient Model, *Journal of Economic Dynamics and Control*, **17**, 233-261.
- CAVAGLIA, S.M.F.G., W.F.C. VERSCHOOR, AND C.C.P. WOLFF, (1994), On the Biasedness of Forward Foreign Exchange Rates: Irrationality or Risk Premia?, *Journal of Business*, **67**, 321-343.
- CECCHETTI, S.G., R.E. CUMBY, AND S. FIGLEWSKI, (1988), Estimation of the Optimal Futures Hedge, *Review of Economics and Statistics*, **70**, 623-630.
- CHAMBERLAIN, G. AND M. ROTHCHILD, (1983), Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets, *Econometrica*, **51**, 1281-1304.
- CHAN, K.S., (1993), Asymptotic Behavior of the Gibbs Sampler, *Journal of the American Statistical Association*, **88**, 320-326.
- CHESNEY, M. AND M. SCOTT, (1989), Pricing European Currency Options: A Comparison of the Modified Black-Scholes Model and a Random Variance Model, *Journal of Financial and Quantitative Analysis*, **24**, 267-284.
- CHO, D.C., C.S. EUN, AND L.W. SENBET, (1986), International Arbitrage Pricing Theory: An Empirical Investigation, *Journal of Finance*, **51**, 313-329.
- CLARK, P.K., (1973), The Subordinated Stochastic Process Model with Finite Variance for Speculative Prices, *Econometrica*, **41**, 135-155.
- CONLEY, T., L.P. HANSEN, E. LUTTMER, AND J. SCHEINKMAN, (1994), Estimating Subordinated Diffusions from Discrete Time Data, unpublished manuscript, University of Chicago.
- CONNOR, G. AND R.A. KORAJCZYK, (1993), A Test for the Number of Factors in an Approximate Factor Model, *Journal of Finance*, **48**, 1263-1291.
- COOPER, I. AND E. KAPLANIS, (1994), Home Bias in Equity Portfolios, Inflation Hedging, and International Capital Market Equilibrium, *Review of Financial Studies*, **7**, 45-60.
- COPELAND, T.E., (1976), A Model of Asset Trading under the Assumption of Sequential Information Arrival, *Journal of Finance*, **31**, 1149-1168.
- DANIELSSON, J., (1994), Stochastic Volatility in Asset Prices: Estimation with Simulated Maximum Likelihood, *Journal of Econometrics*, **61**, forthcoming.
- DANIELSSON, J. AND J.-F. RICHARD, (1993), Accelerated Gaussian Importance Sampler with Application to Dynamic Latent Variable Models, *Journal of Applied Econometrics*, **8**, 153-173.
- DEMPSTER, A.P., N.M. LAIRD, AND D.B. RUBIN, (1977), Maximum Likelihood from Incomplete Data via the EM Algorithm, *Journal of the Royal Statistical Society, Series B*, **39**, 1-38.
- DIEBOLD, F.X., (1988), *Empirical Modeling of Exchange Rate Dynamics*, Springer-Verlag, New York.

- DIEBOLD, F.X. AND M. NERLOVE, (1989), The Dynamics of Exchange Rate Volatility: a Multivariate Latent Factor ARCH Model, *Journal of Applied Econometrics*, **4**, 1-21.
- DOMOWITZ, I. AND C.S. HAKKIO, (1989), Conditional Variance and the Risk Premium in the Foreign Exchange Market, *Journal of International Economics*, **19**, 47-66.
- DROST, F.C., TH. E. NIJMAN, AND B.J.M. WERKER, (1994), Estimation and Testing in Models Containing both Jumps and Conditional Heteroskedasticity, unpublished manuscript, CENTER, Tilburg University.
- DUFFIE, D. AND K.J. SINGLETON, (1993), Simulated Moments Estimation of Markov Models of Asset Prices, *Econometrica*, **61**, 929-952.
- EAKER, M.R. AND D.M. GRANT, (1987), Cross-Hedging Foreign Currency Risk, *Journal of International Money and Finance*, **6**, 85-105.
- ENGLE, R.F., (1982), Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation, *Econometrica*, **50**, 987-1007.
- ENGLE, R.F. AND T. BOLLERSLEV, (1986), Modelling the Persistence of Conditional Variances, *Econometric Reviews*, **5**, 1-50.
- ENGLE, R.F., T. ITO, AND W.-L. LIN, (1990), Meteor Showers or Heat Waves? Heteroskedastic Intra-daily Volatility in the Foreign Exchange Market, *Econometrica*, **58**, 525-542.
- ENGLE, R.F. AND G.G.J. LEE, (1993), A Permanent and Transitory Component Model of Stock Return Volatility, unpublished manuscript, University of California, San Diego.
- ENGLE, R.F. AND G.G.J. LEE, (1994), Estimating Diffusion Models of Stochastic Volatility, unpublished manuscript, University of California, San Diego.
- ENGLE, R.F., D.M. LILIEN, AND R.P. ROBINS, (1987), Estimating Time Varying Risk Premia in the Term Structure, *Econometrica*, **55**, 391-407.
- ENGLE, R.F. AND V.K. NG, (1993), Measuring and Testing the Impact of News on Volatility, *Journal of Finance*, **48**, 1749-1778.
- ENGLE, R.F., V.K. NG, AND M. ROTHCHILD, (1990), Asset Pricing with a FACTOR-ARCH Covariance Structure, *Journal of Econometrics*, **45**, 213-237.
- EPPE, T.W. AND M.L. EPPE, (1976), The Stochastic Dependence in Security Price Changes and Transaction Volumes: Implications for the Mixture-of-Distributions Hypothesis, *Econometrica*, **44**, 305-321.
- EUN, C.S. AND B.G. RESNICK, (1988), The Exchange Rate Uncertainty, Forward Contracts, and International Portfolio Selection, *Journal of Finance*, **43**, 197-215.
- FAMA, E., (1984), Forward and Spot Exchange Rates, *Journal of Monetary Economics*, **14**, 319-338.
- FELLER, W., (1968), *An Introduction to Probability Theory and its Applications*, Wiley, New York.
- FRANKEL, J.A. AND K.A. FROOT, (1987), Using Survey Data to Test Standard Propositions Regarding Exchange Rate Expectations, *American Economic Review*, **77**, 133-153.

- FROOT, K.A., (1993), Currency Hedging over Long Horizons, NBER working paper 4355.
- GALLANT, A.R., D. HSIEH, AND G.E. TAUCHEN, (1994), Estimation of Stochastic Volatility Models with Diagnostics, unpublished manuscript, Duke University.
- GALLANT, A.R., P.E. ROSSI, AND G.E. TAUCHEN, (1992), Stock Prices and Volume, *Review of Financial Studies*, **5**, 199-242.
- GALLANT, A.R., P.E. ROSSI, AND G.E. TAUCHEN, (1993), Nonlinear Dynamic Structures, *Econometrica*, **61**, 871-907.
- GALLANT, A.R. AND G.E. TAUCHEN, (1994), Which Moments to Match?, unpublished manuscript, University of North Carolina.
- GEWEKE, J., (1994), Monte Carlo Simulation and Numerical Integration, working paper 526, Federal Reserve Bank of Minneapolis.
- GHYSELS, E., C. GOURIÉROUX, AND J. JASIAK, (1995), Market Time and Asset Price Movements Theory and Estimation, unpublished manuscript, University of Montreal.
- GHYSELS, E. AND J. JASIAK, (1995), Stochastic Volatility and Time Deformation: an Application of Trading Volume and Leverage Effects, unpublished manuscript, University of Montreal.
- GLEN, J. AND P. JORION, (1993), Currency Hedging for International Portfolios, *Journal of Finance*, **48**, 1865-1886.
- GLOSTEN, L.R., R. JAGANNATHAN, AND D.E. RUNKLE, (1993), On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks, *Journal of Finance*, **48**, 1779-1801.
- GOURIEROUX, C., A. MONFORT, AND E. RENAULT, (1993), Indirect Inference, *Journal of Applied Econometrics*, **8**, 85-118, special issue.
- GRANGER, C.W.J. AND P. NEWBOLD, (1986), *Forecasting Economic Time Series*, Academic Press, San Diego.
- HAMILTON, J.D., (1994), *Time Series Analysis*, Princeton University Press.
- HANSEN, L.P., (1982), Large Sample Properties of Generalized Method of Moments Estimators, *Econometrica*, **50**, 1029-1054.
- HARRIS, L., (1987), Transaction Data Tests of the Mixture of Distributions Hypothesis, *Journal of Financial and Quantitative Analysis*, **22**, 127-141.
- HARVEY, A.C., (1989), *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press, Cambridge.
- HARVEY, A.C., E. RUIZ, AND N. SHEPHARD, (1994), Multivariate Stochastic Variance Models, *Review of Economic Studies*, **61**, 247-264.
- HARVEY, A.C. AND N. SHEPHARD, (1994), Estimation of an Asymmetric Stochastic Volatility Model for Asset Returns, unpublished manuscript, Nuffield College, Oxford.
- HIEMSTRA, C. AND J.D. JONES, (1994), Testing for Linear and Nonlinear Granger Causality in the Stock Price-Volume Relation, *Journal of Finance*, **49**, 1639-1664.
- HODRICK, R.J., (1987), *The Empirical Evidence on the Efficiency of Forward and Futures Foreign Exchange Markets*, Harwood, Chur.

- HODRICK, R.J., (1989), Risk, Uncertainty, and Exchange Rates, *Journal of Monetary Economics*, **23**, 433-459.
- HODRICK, R.J. AND S. SRIVASTAVA, (1986), The Covariation of Risk Premia and Expected Future Spot Exchange Rates, *Journal of International Money and Finance*, **5**, 5-22, special issue.
- HSIEH, D.A., (1989), Modeling Heteroskedasticity in Daily Foreign Exchange Rates, *Journal of Business and Economic Statistics*, **7**, 307-317.
- HULL, J.C. AND A. WHITE, (1987), The Pricing of Options on Assets with Stochastic Volatilities, *Journal of Finance*, **42**, 281-300.
- IKEDA, S., (1991), Arbitrage Asset Pricing under Exchange Risk, *Journal of Finance*, **46**, 447-455.
- INGERSOLL, J.E., (1987), *Theory of Financial Decision Making*, Rowman and Littlefield, Savage, Maryland.
- JACQUIER, E., N.G. POLSON, AND P.E. ROSSI, (1994), Bayesian Analysis of Stochastic Volatility Models, *Journal of Economics and Business Statistics*, **12**, 371-417, with discussion.
- JACQUIER, E., N.G. POLSON, AND P.E. ROSSI, (1995), Models and Priors for Multivariate Stochastic Volatility, unpublished manuscript, Cornell University.
- JARROW, R.A., (1988), *Finance Theory*, Prentice Hall, Englewood Cliffs, New Jersey.
- JOHNSON, H. AND D. SHANNO, (1987), Option Pricing when the Variance is Changing, *Journal of Financial and Quantitative Analysis*, **22**, 143-151.
- JOHNSON, N.L. AND S. KOTZ, (1972), *Distributions in Statistics: Continuous Multivariate Distributions*, Wiley, New York.
- KIM, S. AND N. SHEPHARD, (1994), Stochastic Volatility: Optimal Likelihood Inference and Comparison with ARCH Models, unpublished manuscript, Nuffield College, Oxford.
- KNIGHT, F.H., (1921), *Risk, Uncertainty and Profit*, A.M. Kelley, New York.
- KOEDIJK, C.G., F.G.J.A. NISSEN, P.C. SCHOTMAN, AND C.C.P. WOLFF, (1994), The Dynamics of Short-Term Interest Rate Volatility Reconsidered, unpublished manuscript, University of Limburg.
- KOEDIJK, C.G., M.M.A. SCHAFGANS, AND C.G. DE VRIES, (1990), The Tail Index of Exchange Rate Returns, *Journal of International Economics*, **29**, 93-108.
- KOEDIJK, C.G. AND P.C. SCHOTMAN, (1989), Dominant Movements in Real Exchange Rates, *Journal of International Money and Finance*, **8**, 517-531.
- KRONER, K.F. AND J. SULTAN, (1993), Time-Varying Distributions and Dynamic Hedging with Foreign Currency Futures, *Journal of Financial and Quantitative Analysis*, **28**, 535-551.
- KYLE, A.S., (1985), Continuous Auctions and Insider Trading, *Econometrica*, **53**, 1315-1335.
- LAMOUREUX, C.G. AND W.D. LASTRAPES, (1990), Heteroskedasticity in Stock Return Data: Volume versus Garch Effects, *Journal of Finance*, **45**, 487-498.

- LAMOUREUX, C.G. AND W.D. LASTRAPES, (1994), Endogenous Trading Volume and Momentum in Stock-Return Volatility, *Journal of Business and Economic Statistics*, **12**, 253-260.
- LAUX, P. AND L. NG, (1993), The Sources of GARCH: Empirical Evidence from an Intraday Returns Model Incorporating Systematic and Unique Risks, *Journal of International Money and Finance*, **12**, 543-560.
- LEBARON, B., (1993), The Joint Dynamics and Stability of Stock Prices and Volume, unpublished manuscript, University of Wisconsin.
- LEVICH, R.M. AND L.R. THOMAS, (1993), Internationally Diversified Bond Portfolios: the Merits of Active Currency Risk Management, NBER working paper 4340.
- LEVY, H. AND H.M. MARKOWITZ, (1979), Approximating Expected Utility by a Function of Mean and Variance, *American Economic Review*, **69**, 308-317.
- LEVY, H. AND M. SARNAT, (1970), International Diversification of Investment Portfolios, *American Economic Review*, **60**, 668-675.
- LUCAS, R.E., (1982), Interest Rates and Currency Prices in a Two-country World, *Journal of Monetary Economics*, **10**, 335-360.
- LÜTKEPOHL, H., (1991), *Introduction to Multiple Time Series Analysis*, Springer-Verlag, Berlin.
- MARK, N.C., (1995), Exchange Rates and Fundamentals: Evidence on Long-Horizon Predictability, *American Economic Review*, **85**, 201-218.
- MELINO, A. AND S.M. TURNBULL, (1990), Pricing Foreign Currency Options with Stochastic Volatility, *Journal of Econometrics*, **45**, 239-265.
- MERTON, R.C., (1990), *Continuous-Time Finance*, Blackwell, Cambridge, Massachusetts.
- MILLER, M.H., (1991), *Financial Innovations & Market Volatility*, Blackwell, Cambridge, Massachusetts.
- NELSON, D.B., (1988), *The Time Series Behaviour of Stock Market Volatility and Returns*, PhD thesis, Massachusetts Institute of Technology.
- NELSON, D.B., (1990), ARCH Models as Diffusion Approximations, *Journal of Econometrics*, **45**, 7-38.
- NELSON, D.B., (1991), Conditional Heteroskedasticity in Asset Pricing: a New Approach, *Econometrica*, **59**, 347-370.
- NELSON, D.B. AND C.Q. CAO, (1992), Inequality Constraints in the Univariate GARCH Model, *Journal of Business and Economic Statistics*, **10**, 229-235.
- NELSON, D.B. AND D.P. FOSTER, (1994), Asymptotic Filtering Theory for Univariate ARCH Models, *Econometrica*, **62**, 1-41.
- OWEN, J. AND R. RABINOVITCH, (1983), On the Class of Elliptical Distributions and their Applications to the Theory of Portfolio Choice, *Journal of Finance*, **38**, 745-752.
- PAGAN, A., (1995), The Econometrics of Financial Markets, unpublished manuscript, Australian National University.
- ROGALSKI, R. AND J. VINSO, (1977), Price Level Variations as Predictors of Flexible Exchange Rate Changes, *Journal of International Business Studies*, **5**, 71-81.

- ROSS, S.A., (1976), The Arbitrage Theory of Capital Asset Pricing, *Journal of Economic Theory*, **13**, 341–360.
- ROTHSCHILD, M. AND J. STIGLITZ, (1970), Increasing Risk I: A Definition, *Journal of Economic Theory*, **2**, 225–243.
- SCHOTMAN, P.C., (1994), Priors for the AR(1) Model: Parameterization Issues and Time Series Considerations, *Econometric Theory*, **10**, 579–595.
- SERCU, P., (1980), A Generalization of the International Asset Pricing Model, *Revue de l'Association Française de Finance*, **1**, 91–135.
- SHEPHARD, N., (1994), Partial Non-Gaussian State Space, *Biometrika*, **81**, 115–131.
- SHIMKO, D.C., (1992), *Finance in Continuous Time*, Kolb, Miami.
- SIMS, C.A., (1980), Macroeconomics and Reality, *Econometrica*, **48**, 1–48.
- SMIRLOCK, M. AND L. STARKS, (1988), An Empirical Analysis of the Stock Price-Volume Relationship, *Journal of Banking and Finance*, **12**, 31–41.
- SOLNIK, B., (1974), An Equilibrium Model of the International Capital Market, *Journal of Economic Theory*, **8**, 500–524.
- SOLNIK, B., (1983), International Arbitrage Pricing Theory, *Journal of Finance*, **38**, 449–457.
- SOLNIK, B., (1988), *International Investments*, Addison-Wesley, Reading, Massachusetts.
- SOLNIK, B., (1993), Currency Hedging and Siegel's Paradox: On Black's Universal Hedging Rule, *Review of International Economics*, **34**, 180–187.
- STULZ, R., (1981), A Model of International Asset Pricing, *Journal of the Financial Economics*, **9**, 383–406.
- TANNER, M.A., (1991), *Tools for Statistical Inference: Observed Data and Data Augmentation Methods*, Lecture Notes in Statistics 67, Springer-Verlag, Heidelberg.
- TAUCHEN, G.E. AND M. PITTS, (1983), The Price Variability-Volume Relationship on Speculative Markets, *Econometrica*, **51**, 485–505.
- TAUCHEN, G.E., H. ZHANG, AND M. LIU, (1993), Volume, Volatility and Leverage: A Dynamic Analysis, unpublished manuscript, Duke University.
- TAYLOR, S.J., (1986), *Modelling Financial Time Series*, Wiley, New York.
- TAYLOR, S.J., (1994), Modeling Stochastic Volatility: a Review and Comparative Study, *Mathematical Finance*, **4**, 183–204.
- TIERNEY, L., (1991), Exploring Posterior Distributions Using Markov Chains, unpublished manuscript, University of Minnesota.
- TUCKER, A.L. AND L. POND, (1988), The Probability Distribution of Foreign Exchange Price Changes: Tests of Candidate Processes, *Review of Economics and Statistics*, **70**, 638–647.
- VRIES, C.G. DE, (1992), Stylized Facts of Nominal Exchange Rate Returns, unpublished manuscript, Erasmus University Rotterdam.

Nederlandse samenvatting

Volatiliteit is een belangrijk begrip in de financiële economie. Het is nauw verbonden met het risico dat op financiële titels gelopen wordt. Om goede investeringsbeslissingen te kunnen nemen is het dus belangrijk om volatiliteit te meten. Bijvoorbeeld bij het samenstellen van een beleggingsportefeuille zal een belegger een afweging moeten maken tussen de verwachte opbrengst en de mate van risico. Een belegger zal zijn portefeuillesamenstelling willen veranderen als gevolg van veranderende marktomstandigheden. In dat geval zou hij een model kunnen gebruiken dat beschrijft hoe het verwacht rendement en de volatiliteit door de tijd veranderen. In andere gevallen heeft een belegger een voorspelling nodig voor de toekomstige volatiliteit. Dit is bijvoorbeeld het geval wanneer een belegger de waarde van een optie wenst te bepalen. Om de voorspelling te kunnen doen, zou hij gebruik kunnen maken van een model voor volatiliteit. Modellen voor volatiliteit kunnen ook van groot nut zijn bij het bepalen van strategieën om (gedeelten van) het risico op een beleggingsportefeuille af te dekken. In dit proefschrift hebben wij verschillende manieren onderzocht om volatiliteit te modelleren. In de empirische financiële literatuur is een groot aantal modellen voorgesteld. Wij hebben ons geconcentreerd op een aantal van deze modellen. Het accent bij de analyse van de door ons gekozen modellen ligt op de empirische toetsing.

Volatiliteit krijgt pas een praktische betekenis als wij in staat zijn om het te meten. In hoofdstuk 2 presenteren wij een algemeen raamwerk voor het modelleren van volatiliteit. Dit raamwerk is gebaseerd op de meetbaarheid van stochastische processen. Wij hebben een informatiestructuur gespecificeerd op basis waarvan elk volatiliteitsproces geassocieerd kan worden. Twee populaire klassen van volatiliteitsmodellen die in de literatuur zijn voorgesteld zijn de AutoRegressieve Conditionele Heteroskasticiteit (ARCH) modellen en de Stochastische Volatiliteit (SV) modellen. ARCH modellen worden gekarakteriseerd door het feit dat het volatiliteitsproces dat ze beschrijven gestuurd wordt door informatie uit het verleden. De informatieverzameling waarop SV modellen zijn gebaseerd bevat naast informatie uit het verleden ook informatie omtrent huidige stochastische processen. SV modellen hebben een lange historie in de theoretische financiering. Modellen in continue

tijd, die het gedrag van bijvoorbeeld, financiële titels of portefeuille selectie beschrijven, bevatten allemaal een contemporaine stochastische component. Wij kunnen een model in continue tijd discreet maken door de tijdsas in discrete stukken te verdelen. Het blijkt dan dat een SV model een natuurlijke kandidaat is om het volatiliteitsproces te beschrijven. Er bestaan ook ARCH modellen die een exacte discrete tijdsbenadering zijn voor financiële processen in continue tijd.

Het schatten van volatiliteitsprocessen is een belangrijk onderzoeksterrein binnen de empirische financiering. Het blijkt dat het schatten van SV modellen gecompliceerder is dan het schatten van ARCH modellen. Omdat de informatieverzameling van ARCH modellen slechts informatie uit het verleden bevat, kan de methode van Maximale Aannemelijkheid (MA) eenvoudig worden toegepast. Het volatiliteitsproces in een SV model is latent, en diengevolge kan MA niet worden gebruikt. In hoofdstuk 3 hebben wij een aantal schattingstechnieken voor SV modellen besproken en toegepast. De analytische vorm van het schattingsprobleem voor een SV model wordt gekarakteriseerd door een multidimensionale integraal. De totale dimensie is gelijk aan het aantal waarnemingen in de steekproef. In de toepassingen die wij in dit hoofdstuk presenteren hebben wij gekozen voor een schattings-techniek die het volatiliteitsmodel transformeert naar een conditionele lineaire en Gaussische toestands-ruimte vorm. Met behulp van het Kalman filter kunnen schattingen voor de onderliggende volatiliteit bepaald worden. De vorm van de meetfout in de meetvergelijking van de toestands-ruimte wordt bepaald door een *mixture* van normale verdelingen. Het gemiddelde en de variantie van elke *mixture* zijn vrije parameters. Schattingen van de hyperparameters in het model zijn bepaald met behulp van simulatiemethoden gebaseerd op een Gibbs algoritme. Wij hebben deze schattingstechniek toegepast zowel in een klassieke context met een gesimuleerd EM algoritme als in een Bayesiaanse context. Ter illustratie van beide technieken hebben wij ze gebruikt in een analyse naar de vorm van wisselkoersnieuws. De wisselkoersreeksen die wij gebruikt hebben zijn gebaseerd op 4 valuta's: Amerikaanse dollar, Japanse yen, Duitse mark en Engelse pond. De waarnemingen zijn verzameld op wekelijkse basis en beslaan de periode van 1973 tot 1994. De geschatte volatiliteitsprocessen verschillen aanzienlijk wanneer een andere schattingsmethode gebruikt wordt. Voor SV modellen kan een spreidingsmaatstaf voor de volatilititeiten uitgerekend worden. Bij berekening blijkt dat de standaarddeviaties op de volatiliteit aanzienlijk zijn. Dit wordt gedeeltelijk veroorzaakt door de lage signaal-ruis verhouding in de SV modellen. De resultaten impliceren dat voorzichtigheid moet worden betracht bij het gebruik van volatiliteit in beslissingsprocessen, bijvoorbeeld in het waarderen van optiecontracten.

De diagnostische toetsen van de univariate SV modellen in hoofdstuk 3 toonden aan dat het univariate SV model met een AR(1) structuur voor de volatiliteit waarschijnlijk lijdt aan misspecificatie. Verder blijkt dat er sprake is van significante verbanden tussen de verschil-

lende volatiliteitsreeksen onderling. Dit motiveert het ontwikkelen van een multivariaat model voor wisselkoersvolatiliteit in hoofdstuk 4. De modellen die tot nu toe zijn gepresenteerd in de literatuur zijn ontwikkeld voor het beschrijven van een wisselkoerssysteem met één enkele valuta als numeraire. Als de numeraire valuta verandert, verandert de vorm (en de interpretatie) van het model. Wij ontwikkelen een nieuw model, dat onafhankelijk is van de valuta, die als numeraire fungeert. Hiertoe veronderstellen wij dat nieuws in wisselkoersen verdeeld kan worden over de verschillende afzonderlijke valuta's. Met andere woorden, wij maken gebruik van een factorstructuur voor een systeem van wisselkoersen waarin de factoren gelijkgesteld worden aan het nieuws dat slechts aan één bepaalde valuta gerelateerd is. In een analyse waarin wij de steekproef opdelen in verschillende perioden blijkt dat de varianties van de nieuwscomponenten in de tijd veranderen. Dientengevolge hebben wij een multivariaat SV model ontwikkeld waarin de factorstructuur voor wisselkoersen behouden blijft. Als gevolg van het gebruik van de factorstructuur heeft het model veel minder parameters nodig dan het geval zou zijn geweest in multivariate ARCH modellen uit de literatuur. Na schatting van ons model vinden wij dat er inderdaad bewijs kan worden gevonden voor de stelling dat de volatiliteiten van de vier belangrijkste valuta's invloed op elkaar kunnen uitoefenen. De volatiliteit in de Amerikaanse dollar wordt snel verwerkt in de volatiliteit van de andere valuta's. Dit is in tegenstelling met de andere valuta's voor welke deze snelheid veel lager is. De Duitse mark blijkt de meest stabiele valuta te zijn geweest over de laatste 20 jaar. Als laatste hebben wij in dit hoofdstuk gekeken naar de implicaties van ons wisselkoerssysteem op risico premies. Wij vinden dat de geschatte risico premies groter zijn dan eerder werd gerapporteerd in de literatuur. Dit wordt voor een belangrijk deel verklaard door de schattingsfout op de volatiliteitsreeksen.

Zowel ARCH als SV modellen zijn statistische modellen. Er is nog geen geschikte economische theorie voorhanden die het bestaan van dit soort modellen rechtvaardigt. Echter, op empirische grond zijn beide modellen redelijk succesvol in het beschrijven van financiële tijdreeksen. Een populaire theorie voor het bestaan van volatiliteit op financiële markten is de zogenaamde *Mixture of Distributions* theorie. Het stochastische proces van een financiële titel wordt gestuurd door een onderliggend proces dat op zijn beurt weer gestuurd wordt door informatie impulsen. Een variabele die vaak in verband wordt gebracht met deze informatie is de verhandelde hoeveelheid van de desbetreffende financiële titel. In hoofdstuk 5 presenteren wij enkele empirische analyses naar het verband tussen handelsvolume en volatiliteit. Wij maken voornamelijk gebruik van het SV model. Een belangrijke veronderstelling die wij maken is dat het handelsvolume en de volatiliteit op hetzelfde tijdstip bepaald worden. Dit impliceert dat het handelsvolume als een schaafactor voor de volatiliteit gezien kan worden. Omgekeerd kan volatiliteit een schaafactor voor het handelsvolume zijn. De schattingsresultaten bevestigen dat het handelsvolume een significante schaafactor is. De data die wij in de analyse gebruiken bestaat uit dagelijkse

rendementen en handelsvolumes van een vijftal Amerikaanse aandelen en de Standard & Poors 500 aandelenindex. De steekproefperiode beslaat de periode van januari 1967 tot en met december 1987. In de empirische literatuur met betrekking tot het beschrijven van aandelenrendementen is vrij veel aandacht besteed aan het bestaan van een hefboomeffect tussen rendementen en volatiliteit. Positieve en negatieve innovaties kunnen een afwijkend gevolg hebben voor de volatiliteit. Wij onderzoeken of er aanwijzingen bestaan voor het hefboomeffect in onze steekproef. Wij vinden dat voor de index er inderdaad sprake is van een significant hefboomeffect, maar voor de individuele aandelen is het niet duidelijk. Het blijkt dat het hefboomeffect voornamelijk wordt veroorzaakt door een aantal extreme waarnemingen in de steekproef. In de SV modellen hebben wij de mogelijkheid opengelaten voor een correlatie tussen de storingstermen in de toestands-ruimte formulering. Het blijkt dat het toelaten van het hefboomeffect niet veel invloed heeft op de relatie tussen handelsvolume en volatiliteit.

In hoofdstuk 6 en hoofdstuk 7 hebben wij twee belangrijke onderwerpen in de internationale financiering onderzocht. In hoofdstuk 6 hebben wij gekeken naar het afdekken van wisselkoersrisico in een internationale portefeuille. Hoofdstuk 7 behandelt het spreiden van vermogen met behulp van een internationale versie van de *Arbitrage Pricing Theory* (APT). Het afdekken van wisselkoersrisico heeft grote aandacht gekregen in de literatuur. Veel minder aandacht krijgen strategieën die expliciet rekening houden met de empirisch gevonden onconditionele leptokurtosis ("dikke staarten") in financiële tijdreeksen. Wij ontwikkelen een model waarin een belegger zijn verwachte nut maximaliseert. De keuzevariabele in dit probleem is een ratio die aangeeft hoeveel wisselkoersrisico de belegger wenst af te dekken. Deze ratio wordt de *hedge* ratio genoemd en is gedefinieerd als het percentage van de uitstaande buitenlandse investering dat afgedekt wordt met behulp van termijncontracten. Van belang voor het bepalen van de *hedge* ratio is het rendement op een portefeuille die in ons geval bestaat uit beleggingen in een binnenlandse en een buitenlandse aandelenindex. Deze portefeuille wordt aangevuld met een aantal termijncontracten. In dit hoofdstuk beschouwen wij de keuze omtrent de fractie die een belegger wenst te beleggen in het buitenland als gegeven. De belegger dient nu het verwachte nut van de totale portefeuille te maximaliseren door een geschikte *hedge* ratio te kiezen. De resulterende ratio is een functie van de momenten van de onderliggende verdeling van de rendementen op de wisselkoers, de aandelenindices en het renteverskil. Wij vatten de gezamenlijke verdeling van deze rendementen samen door direct een verdeling te specificeren voor het rendement op de gehele portefeuille. Voor deze verdeling kiezen wij een Student t verdeling. Deze verdeling heeft dikkere staarten dan de normale verdeling. Voor een reeks van *hedge* ratio's kunnen nu de parameters van de Student t geschat worden en vervolgens kan het verwachte nut uitgerekend worden. De dikstaartigheid die wij vinden in de data heeft slechts invloed op de optimale *hedge* ratios wanneer wij de nutsfunctie zodanig aanpassen dat er (negatief)

nut wordt toegekend aan de leptokurtosis.

Een goed model voor het waarderen van beleggingsobjecten in een internationale omgeving moet voldoen aan de eis dat de implicaties niet veranderen als de valuta waarin de rendementen uitgedrukt worden, verandert. Als aan deze eis niet voldaan is, kunnen beleggers risicoloze winsten behalen door de numeraire valuta te veranderen. Het Internationale APT (IAPT) is het meest onderzochte model in de literatuur. Om de invariantie ten opzichte van de numeraire valuta te waarborgen zijn echter sterke aannames nodig. Wij laten zien dat een IAPT model geconstrueerd kan worden waarin deze restricties overbodig blijken. Het model is gebaseerd op het feit dat factoren in een factor model, zoals het IAPT, kunnen worden benaderd door een principale componenten analyse. Doordat de covariantiestructuur behouden blijft in een principale componenten analyse, is het mogelijk om een zodanige rotatie van de factoren te verkrijgen dat deze onafhankelijk worden van de valuta waarin de wisselkoersen worden uitgedrukt.

Dit proefschrift heeft een aantal nieuwe inzichten gegeven omtrent het modelleren van volatiliteit. Volatiliteit blijkt moeilijk te meten, zowel vanuit een statistisch als ook vanuit een empirisch oogpunt. Resultaten met een stochastisch volatiliteitsmodel laten zien dat de signaal-ruis ratio laag is. Verder blijkt dat geschatte volatiliteitsprocessen hoge standaarddeviaties hebben. De oorsprong van volatiliteit in financiële markten is allerm minst duidelijk. Behalve het perfectioneren van statistische methoden om volatiliteit te meten, is het ook belangrijk om de factoren die volatiliteit veroorzaken te onderzoeken. Eén manier om dat te doen is het analyseren van microstructuren van financiële markten. Het recentelijk beschikbaar komen van data met zeer hoge frequentie kan een belangrijke impuls geven aan het empirische onderzoek. Een andere manier om de oorsprong van volatiliteit te onderzoeken is om gebruik te maken van experimentele markten. Het voordeel is dat bepaalde marktsituaties gecontroleerd kunnen worden door de onderzoeker. De toekomst zal moeten uitwijzen of deze methoden succesvol zullen zijn.

Index

- Abramowitz, 99
Adler, 120, 121, 139, 140, 143
Admati, 10, 85
Akgiray, 128
Andersen, 5, 10, 12, 13, 18, 19, 26, 27, 46,
85, 86
Anderson, 120, 129

Baillie, 20, 24, 67, 78, 129
Basilevsky, 143, 145
Bauer, 86, 88, 106
Bekaert, 78
Bernstein, 1
Black, 9, 12, 13, 93, 120
Bollerslev, 6, 9–11, 19, 20, 24, 60, 67, 78,
128
Booth, 128
Boothe, 62, 120, 128
Brenner, 86
Brock, 10

Campbell, 85, 93
Canova, 31
Cao, 11
Cavaglia, 82
Cecchetti, 129
Chamberlain, 139, 144
Chan, 28, 29
Chesney, 13
Cho, 143
Chou, 10, 11, 78
Clark, 9, 85, 87
Conley, 22
Conner, 142
Cooper, 121
Copeland, 85
Cumby, 129

Danielsson, 30
Danthine, 120, 129

De Vries, 31, 120, 135
Dempster, 29, 33
Diebold, 67, 92, 120
Domowitz, 10, 60, 62, 78
Drost, 14, 22
Duffie, 26
Dumas, 120, 121, 139, 140, 143

Eaker, 120
Engle, 6, 9–11, 15, 19, 20, 22, 24, 51, 67,
68, 74, 86, 93, 94, 97
Epps, M.L., 85
Epps, T.W., 85
Eun, 123, 127, 143

Fama, 84
Feller, 2
Figlewski, 129
Foster, 14
Frankel, 78, 82
Froot, 78, 82, 119

Gallant, 20, 26, 27, 43, 85, 86, 89, 95, 97
Geweke, 30
Ghysels, 22, 85
Glassman, 62, 120, 128
Glen, 119, 121
Glosten, 93
Gourieroux, 22, 27
Granger, 123
Grant, 120
Grossman, 85

Hakkio, 60, 62, 78
Hamilton, 107
Hansen, 22, 26
Harjes, 86
Harris, 85
Harvey, 23, 29, 60, 67, 84, 97, 100, 102
Hentschel, 93

- Hiemstra, 108
 Hodrick, 62, 78, 84
 Hsieh, 20, 26, 27, 43, 67, 128
 Hull, 13

 Ikeda, 139, 141–143, 146, 147
 Ingersoll, 3, 4, 134
 Ito, 74

 Jacquier, 19–21, 27, 51, 88, 97, 99
 Jagannathan, 93
 Jansen, 135
 Jasiak, 22, 85
 Johnson, 13, 133
 Jones, 108
 Jorion, 119, 121

 Kaplanis, 121
 Kim, 20, 28–30, 36
 Kleidon, 10
 Knight, 2, 3
 Koedijk, 86, 120, 135, 139, 144–146
 Korajczyk, 142
 Kotz, 133
 Kroner, 10, 11, 78, 86, 129
 Kyle, 85

 Lütkepohl, 107, 108
 Laird, 29, 33
 Lamoureux, 85–89, 106
 Lastrapes, 85–89, 106
 Laux, 85, 86
 LeBaron, 85, 107, 108
 Lee, 22, 51, 86, 97
 Levich, 119
 Levy, 5, 121
 Lilien, 15
 Lin, 74
 Liu, 85, 86
 Lucas, 78, 84
 Luttmmer, 22

 Mark, 31
 Markowitz, 1, 5
 Marshall, 78
 Melino, 13, 26, 27
 Merton, 13
 Miller, 5

 Monfort, 22, 27
 Myers, 129

 Nelson, 6, 9–11, 14, 15, 19, 60, 128
 Nerlove, 67
 Newbold, 123
 Ng, L., 85, 86
 Ng, V.K., 68, 93, 94
 Nieuwland, 86, 88, 106
 Nijman, 14, 22
 Nissen, 86

 Owen, 134

 Pagan, 10, 16, 20
 Pfeiderer, 10, 85
 Pitts, 10, 85, 87
 Polson, 19–21, 27, 51, 88, 97, 99
 Pond, 128
 Prasad, 120

 Rabinovitch, 134
 Renault, 22, 27
 Resnick, 123, 127
 Richard, 30
 Robins, 15
 Rogalski, 128
 Ross, 6, 140
 Rossi, 19–21, 27, 51, 85, 86, 88, 89, 95, 97, 99
 Rothschild, 4, 68, 139, 144
 Roy, 1
 Rubin, 29, 33
 Ruiz, 23, 60, 67, 84, 99, 102
 Runkle, 93

 Sørensen, 19, 26, 27, 46
 Sarnat, 121
 Schafgans, 120
 Scheinkman, 22
 Scholes, 9, 13
 Schotman, 39, 86, 139, 144, 146
 Scott, 13
 Senbet, 143
 Sercu, 120
 Shanno, 13
 Shephard, 20, 23, 28–31, 36, 54, 60, 67, 84, 97, 100, 102

Shimko, 14
Sims, 107
Singleton, 26
Smirlock, 85
Solnik, 120, 126, 139, 141, 143, 146, 147
Srivastava, 84
Starks, 85
Stegun, 99
Stiglitz, 4
Stulz, 120
Sultan, 129

Tanner, 29, 33
Tauchen, 10, 20, 26, 27, 43, 85–87, 89, 95,
97
Taylor, 26, 67
Thomas, 119
Tierney, 29
Tucker, 128
Turnbull, 13, 26, 27

Verschoor, 82
Vins0, 128

Wang, 85
Werker, 14, 22
White, 13
Wolff, 82, 86
Wooldridge, 67

Zhang, 85, 86

Curriculum Vitae

Ronald J. Mahieu is op 7 mei 1968 geboren te Rotterdam. In 1986 slaagde hij voor het VWO eindexamen aan de C.S.G. Comenius te Capelle aan den IJssel. Vervolgens is hij Econometrie gaan studeren aan de Erasmus Universiteit in Rotterdam. In 1991 behaalde hij zijn doctoraaldiploma. In april 1991 werd hij assistent in opleiding bij de sectie Financiering van de Rijksuniversiteit Limburg. Hij werkte daar aan verschillende onderzoeksprojecten, waarvan dit proefschrift het resultaat is. Van januari 1992 tot en met mei 1992 was hij *Visiting Student* aan de *Woodrow Wilson School of International and Public Affairs* van de universiteit van Princeton in de Verenigde Staten. Van eind 1992 tot en met 1993 werkte hij op part-time basis bij de staf Strategische Beleidsontwikkeling Beleggingen van het Algemeen Burgerlijk Pensioenfonds. In 1993 behaalde hij het diploma van het Netwerk Algemene en Kwantitatieve Economie. Vanaf mei 1995 is hij werkzaam bij Ortec Consultants B.V. te Gouda. Daarnaast is hij als *Postdoctoral Fellow* verbonden aan de Rijksuniversiteit Limburg.

Ronald J. Mahieu, born May 7th 1968 in Rotterdam, studied Econometrics at the Erasmus University of Rotterdam from 1986 until 1991. He received his Master's degree in Econometrics in April 1991. In May 1991 he joined the Finance Department of the University of Limburg in Maastricht. There he started working on this doctoral thesis. During the spring semester of 1992 he was a Visiting Student at the Woodrow Wilson School of International and Public Affairs of Princeton University, USA. He worked for the Strategic Investment Policy Group of the *Algemeen Burgerlijk Pensioenfonds* from 1992 to 1993. Starting May 1995 he is working at Ortec Consultants B.V. in Gouda. At the same time he is employed as a Postdoctoral Fellow at the University of Limburg in Maastricht.